5TH SOUND AND MUSIC COMPUTING CONFERENCE

BERLIN - GERMANY

SMC08

JULY 31ST - AUGUST 3RD 2008

SOUND IN SPACE SPACE IN SOUND

PROCEEDINGS

Edited by Martin Supper, Stefan Weinzierl

Organized by
German Association of Electroacoustic Music (DEGEM)
in collaboration with the Audio Communication Group,
Technische Universität Berlin/Germany

The conference will take place at Technische Universität

SMC 08
5th Sound and Music Computing Conference

sound in space – space in sound

July 31st - August 3rd, 2008, Berlin, Germany

Proceedings

Edited by
Martin Supper
Stefan Weinzierl

Organized by
German Association of Electroacoustic Music (DEGEM)
in collaboration with the
Audio Communication Group, Technische Universität Berlin/Germany.

The conference will take place at Technische Universität

supported by

DEGEM

DEUTSCHER MUSIKRAT

ZKM
Zentrum für Kunst und
Medientechnologie Karlsruhe

SENNHEISER

TU berlin

ohrenstrand.net

Netzwerk
Neue Musik

ein Förderprojekt der
KULTURSTIFTUNG
DES
BUNDES

SMC 08
5th Sound and Music Computing Conference

sound in space – space in sound
July 31st - August 3rd, 2008, Berlin, Germany

Conference Committee
Michael Harenberg
Folkmar Hein
Martin Supper
Stefan Weinzierl

Scientific Chair
Stefan Weinzierl

Concerts and Installations
Folkmar Hein

Proceedings Editors
Martin Supper
Stefan Weinzierl

Local organizing committee
Christian Dietz
Florian Goltz
Eckehard Güther
Folkmar Hein
Thilo Koch
Andreas Lehmann
Hans-Jochim Mond
Zora Schärer
Niklas Schminke
Frank Schultz
Wilm Thoben

Public Relations
Yvonne Stingel

Online Paper Submission System
Paul Modler

Music Committee
Björn Gottstein (Germany)
Michael Harenberg (Germany)
Folkmar Hein (Germany)
Fernando Lopez-Lescano (USA)
Martin Supper (Germany)

Scientific Review Committee
Marije Baalman
(Concordia University, Canada)
Roger Dannenberg
(Carnegie Mellon University, USA)
Dominique Fober
(GRAME, France)
Stefan Kersten
(University Pompeu Fabra, Spain)
Fernando Lopez-Lescano
(CCRMA, USA)
Eduardo Reck Miranda
(University of Plymouth, UK)
Emmanuel Saint-James
(Université Paris IV, France)
Norbert Schnell
(IRCAM, Paris)
Martin Supper
(University of the Arts, Berlin)
Stefan Weinzierl
(Technische Universität, Berlin)

Conference Secretariat
Doris Grasse

Webmaster
Andreas Lehmann (TU Berlin)
Elmar Farchmin (HfG Karlsruhe)

Graphic-Design
Anne-Claire Martin

Foreword/Presentation

Ladies and Gentlemen,
Dear Colleagues and Friends !

On behalf of the German Association of Electroacoustic Music (DEGEM), I would like to welcome you to the 5th SMC 2008 in Berlin. Berlin as one of the major cultural sites of Europe has always been and still is today a center of musical experimentation and intercultural encounters. Therefore, we are particularly delighted to be here. We are also very happy about the cooperation with the Audio Communication Group of the TU Berlin, who has the largest permanently installed wave field synthesis system worldwide, an installation unique for exploring newly and enhanced sound spaces.

With our focus on the space of music as well as on musical spaces, which was the inspiration to our title “Space in Sound – Sound in Space", we have chosen an exigent subject theoretically and aesthetically very much up-to-date. Moreover, we are lucky enough to experience and explore the potential of a unique constellation of sound spaces including the Acousmonium (GRM Paris), the Sound Dome (a la ZKM, Karlsruhe), and the Wave Field Synthesis (TU Berlin) from a historical perspective, as well as with an eye to the future.

Every era develops its specific, culturally defined awareness of space as well as forms of its aesthetic reification. In music, we can trace a development from an architectural place of sound to the symbolical space of formal and structural projections and finally to the imaginitiv, musically immanent space of compositional fantasy.

From thereon the actual space can be functionalised musically as one possibility. It can, however, also be opened to and expanded by technical spaces. These, from digital simulations to virtualities, enable both universal manipulation and boundless scaling. Thus, the conception of an "acoustic cyberspace" (Harenberg, 2003) which is technically primarily conveyed by time modes becomes constitutive for new aesthetical conceptions of form as well as for the generation and manipultion of sound.

The historical circle is opening nowadays, as the early form and thus structure giving functions of space of the Renaissance as technical and structural augmentations of compositional and formal principles are finding a new "language of sound and form" through the acoustic "Pearly Gates of Cyberspace" (Wertheim, 2000), which technological fundamentals, backgrounds, fantasies and current applications the SMC 08 in Berlin wants to fathom and investigate.

I would like to thank the Audio Communication Group of the TU Berlin, and especially Stefan Weinzierl, for the organization of this year's SMC as a partner of the DEGEM. The DEGEM is a small association and would not have been able to organise an international event like this on its own. Therefore we depend on the organisational and conceptional support of such a competent and potent partner.

I also want to thank Folkmar Hein, without whose commitment the tight cooperation with the festival "INVENTIONEN" would not have been possible. A unique concert and exhibition program resulted from the affinity of both events in terms of content as well as organisation, and builds a more than interesting framework for this year's SMC.

Last but not least I would like to sincerly thank all of those who - from the juries, the technical installations, to the live broadcast of SMC 08 in the DEGEM WebRadio@ZKM - made this SMC in Berlin possible. They have all contributed also to the further successful development of the SMC.

And so, I wish all of us wonderful and demanding days here in Berlin, plenty of interesting music, a lot of exiting lectures, discussions and encounters.

Michael Harenberg

DEGEM

Steering Commitee

France:
Myriam Desainte-Catherine, Labri, University of Bordeaux
Dominique Fober, Grame, Lyon GMEM - Marseille
Charles Bascou, GMEM, Marseille

Germany:
Martin Supper, University of the Arts, Berlin
Michael Harenberg, DEGEM, Karlsruhe

Greece:
Anastasia Georgaki, National and Capodistrian University of Athens - Music Department
Ioannis Zannos, Ionian University

Italy:
Davide Rocchesso, University of Verona
Riccardo Dapelo, Conservatorio di Sassari

Sound and Music Computing (SMC) is supervised jointly by AFIM (Association Française d'Informatique Musicale), AIMI (Associazione Italiana di Informatica Musicale), DEGEM (Deutsche Gesellschaft für Elektroakustische Musik), and HACI (Hellenic Association of Music Informatics).

SMC Concerts

Friday, 1.8. - TU Berlin WFS-hall 0104

7 pm SMC08 – Presentation I (call for music)

Erik Nyström *Multiverse* 2008 11’07 Acousmonium
Volker Hennes *The Maelstrom Method* Klangdom (inner quadro) 7’07
Sam Salem *They Sing for Themselves* Klangdom + WFS 6’54
Douglas Henderson *The 103rd Thing and the 104th Thing (of 10.000)*
2003-06 WFS + Klangdom + Acousmonium 11’35
Georg Dennis *Electric Sheep* 2008 Klangdom 5’04
Jef Chippewa *DUO* 1997-98 Acousmonium 2’41
Javier Alejandro Garavaglia *Pathétique* 2006/2007 Klangdom 15’30
Martin Bedard *Excavations* 2008 Acousmonium 10’

10.30 pm SMC08– Presentation II (call for music)

Thanos Chrysakis *INSCAPES* 11-10 2005 Klangdom 10’20
John Ritz *In the Very Eye of Night* Acousmonium
Yutaka Makino *Ephemera* 2008 WFS + Klangdom + Acousmonium 10’10
Daniel Blinkhorn J*eu fabriqué* Acousmonium
Ka Ho Cheung *FishyBahn* 2008 Klangdom 9’48
Pei Yu Shi *Fall aus der Zeit* … 2006 WFS 10’
Manuella Blackburn *Origami* 2007 Acousmonium Stereo 5’10
Ioannis Kalantzis *Parastaseis A B C D* 2003-2006 any system 7’
Annette Vande Gorne *Yawar Fiesta* (opera), 2. part Acousmonium + Klangdom 10’

Location:

TU Berlin
Main building
Room H 0104
Straße des 17. Juni 135
Charlottenburg
U2 - Ernst-Reuter-Platz
S - Tiergarten

Erik Nyström *Multiverse* 2008
Discontinuous connections and connected discontinuities lie at the heart of this piece. These aspects are reflected both in the choice of sound material and the way the piece is structured. Percussive singularities are presented and approached on several parallel strata, from "big" obvious gestural events, across rhythmic pulsations, towards more abstract textures and drones. The boundaries are blurred as a synthesis takes place within the network of connections that constitute the composition, weaving an irregular fabric of space and time – full of knots and holes – where music emerges in a gravitational flux.

The term multiverse is used in cosmological science and describes a constellation of universes, where phase transitions such as chaotic inflations and big bangs spawn new regions in space and time.
The piece was premiered at the ElectriCity festival in London in April 2008 and was shortlisted for the Residence prize in the 35th Bourges International Competition in 2008.

Volker Hennes *The Maelstrom Method* 2007/08
The Maelstrom Method is a quadrophonic acousmatic piece, composed in 2007/08.
Inspired by a short-story by E. A. Poe, the composition follows an almost anecdotic approach to conceive an epic progression. The piece acoustically spotlights spacial elaboration as an imaginary dimension – formed by the use of trajectories and motions – to build immersive structure as undirected radiation.
"The mountain trembled to its very base, and the rock rocked. I threw myself upon my face, and clung to the scant herbage in an excess of nervous agitation.(...) As I felt the sickening sweep of the descent, I had instinctively tightened my hold upon the barrel, and closed my eyes. For some seconds I dared not open them while I expected instant destruction, and wondered that I was not already in my death-struggles with the water." Excerpt from "A Descent Into the Maelstrom" by Edgar Allan Poe.
First performance took place at the Oscar Peterson Concert Hall on the 08. February 2008 within the "EuCuE Concerts" – arranged by the Concordia University Montreal/ Canada, curated by Kevin Austin.

Sam Salem *They Sing for Themselves* 2008
I wanted to explore the themes of surveillance, paranoia and the ubiquitious sinister technologies with which we co-exist. Instead, and rather fortunately, I found loveliness and charm. They Sing for Themselves is an abstraction of our soundscapes, an imaginary landscape for the ear: how we choose to navigate it is up to us. They are always there, singing and sputtering. This piece is part of a larger work that spans media and contexts.

Douglas Henderson *The 103rd Thing and the 104th Thing (of 10.000)* 2003-2006
Begun in the summer of 2001 and still ongoing, The Cycle of the 10,000 Things explores the sculptural potential of multichannel audio as a primary focus for composition. The nature of the “music” in each piece develops from the dictates of sonic holography; the temporal progression follows a physical model similar to the construction of a house. I often use multiplications of simple acts to reveal their underlying and otherworldly qualities. The 103rd Thing, a sensual cataclysm of shattering glass with a slow-motion portrait of the movement of the shards, includes some 700 glass cuts, and the sounds of thousands of microscopic glass particles settling onto a pane. The 104th Thing dissects recordings of 200 cups of coffee made in a dying espresso machine. Originally 6 channel works, these have been re-orchestrated to take advantage of the special properties of the WFS and Klangdom systems.

Georg Dennis *Electric Sheep* 2008
The forefather of modern computer science, Alan Turing, once theorised that were a computer ever to effectively simulate the human mind, it must also inherit its mistakes and failings, its tendency towards error.
Functioning correctly, the modern computer seems almost incapable of such faults; it is a model of unthinking perfection. Perhaps then, any audible 'errors' that computers make and have made throughout their history could be extraneous by-products of their operation: the screech of the metal housing containing the early machines that Turing worked with; the clicks and whirrs of tape players once used as storage media; the modern-day hums and buzzes of the computers that we are all familiar with.

This piece makes use of these sounds. However, an attempt has been made to imbue these inanimate sonic objects with the spark of life whilst still retaining something of a machine-like quality, to give them a sense of intelligence whilst ultimately remaining artificial.
Electric Sheep is dedicated to the memories of Alan Turing and Philip K. Dick, the latter's novel/treatise on 'AI' providing the title of this work.

Jef Chippewa *DUO* 1997-98
was composed in the concordia ea studios during the 1997-8 year. the compositional and perceptual experience of DUO offers the potential for comprehension of musical interactions and appreciation of the correspondence of timbre and articulation types between two radically different worlds of sound production (that of the [alto] saxophone and of the analog synthesizer), instead of a lethargic experience of linear continuity, and establishment of familiarity on a superficial level. thanks to yves charuest for his openness and flexibility as a performer and as a musician.

Javier Alejandro Garavaglia *Pathétique* 2006/2007
The piece works with mainly materials extracted from the first three main chords of the Introduction (Grave) to the 1st Movement of Beethoven's Sonata Op. 13 in C minor, "Pathétique". The chords are: C Minor, its Subdominant (diminished 7th chord) and the VII7 of C Minor (another diminished 7th chord).

The positions used by Beethoven for these chords make the sound of the piano very rich in harmonics. Therefore the piece explores the richness of their spectrum and transforms it accordingly. It can be viewed as an exploration form the side of the listener "into" the sound. This journey has a double aspect, as it refers not only to the new spectral results overall, but also to the distribution of sound in space (Spatialization in 5.1 or 8.0).

The DSP processes are very varied, using different tools like C-Sound, Lisp programming, Audiosculpt, GRM tools, SoundHack, Peak, etc. and involving mainly Time stretching and Pitch shifting, Granulation, Filtering, Phase shifting, Envelope shaping, etc.

Martin Bedard *Excavations* 2008
Commissioned for the Québec City 400th anniversary (1608-2008) celebrations. Excavations is a homage to the history and unique character of Québec City. In the piece, I explore the cohabitation of electroacoustic media and sound culture, which I identify as being the unique sound heritage of a community or area. The composition uses referential sounds, which are recognizable and anchored in reality. These have then been reworked in the studio to transform their anecdotal nature into material that can be presented in musical form. The sounds have been used as symbols, metaphors and indices, here suggesting a narrative approach to the design of the sound phenomenon. Non-referential sounds, created using montage and treatment techniques, have been added to form part of the cohabitation. They punctuate the écriture of the sound into phrasings, take on the role of signals, or have a function that is purely abstract. The title Excavations alludes to the archaeological campaigns in the city.

Thanos Chrysakis *INSCAPES 11-10* 2005
Inscapes is a series of sound based compositions in which my primary concern lies on the timbral affinities and contrasts of the sounds. They have been inspired by Gerard Manley Hopkins' idea of the inscape as «species or individually-distinctive beauty», and in consequence, the exploration of the inherent qualities of different sound matter, for generating specific aural structures, and what I define in my work as aural morphogenese. In addition -after sometime working on them-, the finding of the words : «enter into oneself, that is to discover subversion» by Edmond Jabés, suggested me a very interesting connection between the interiority of the sound, and the interiority of the composer, which both intersect in the act of listening.

John Ritz *In the Very Eye of Night*
The laws of macro- and microcosm are alike. Travel in the interior is as a voyage in outer space: we must in each case burst past the circumference of our surface – enter worlds where the relationship of parts is the sole gravity.

Yutaka Makino *Ephemera*

Daniel Blinkhorn *Jeu fabriqué*
Recollections of industry, fabrication and the mechanical arts provided the foundation for jeu fabriqué…
As a child, the seemingly endless mechanised space of my father's workshop, and all the sonic activity contained within, became augmented through my imagination.

Each time I would visit the workshop, an expansive spectral palate seemed to unfold, where work tools, the shapes and sounds of these tools and the spaces surrounding them provided vehicles of discovery… Tools became toys, articulating the imagined.

The workshop was a place of motion, industry and invention…Positioned within dense foliage (and home to much birdlife), I was not surprised to find that my recollections of the workshop were entwined with images of its surrounds.

Throughout the work, intersecting patterns, gestures and spaces are presented to the listener, modulating between the abstract and the concrete…Images are plotted, and sights and sounds unravel as recollection and chimera become fused.

The material heard in the piece was generated from recordings of toy tools, real tools, imagined spaces and real spaces, all of which attempt to typify some of the sonorities and imagery found within an environment of fabrication, invention and imagination.

Ka Ho Cheung *FishyBahn* 2008
Fishy Bahn is originally written for eight loudspeakers. Sound materials are mainly based on urban rails in Berlin (S-Bahn and U-Bahn), with the aid of ICST ambisonics tools in spatialization. As shown in picture below, I imagine the trains are like fishes. They swim and reach every corner of the city – the rich and the poor, the communist and the capitalist monuments. I see their red tails swinging, their motions are smooth and speedy. Pitches from the engines shifting up-and-down, I feel tension-and-release. Freshly wild during the rush hours, they scream along the silver rails. Sleepy during the night, they swipe their dreamy lights through the dark…

Pei Yu Shi *Fall, aus der Zeit …*2006
A friend of mine wanted to realise a dance project about Ingeborg Bachmann and she sent me one of her poems and texts to be set to music. My aim was to get an insight into the inner world of the poet. I was asking for her moods when she wrote the poems. Then I began to try to set a poem by Ingeborg Bachmann musically and to hint at a mental and emotional change on different levels through the process of scenic narrative.

Poem:

Fall ab, Herz
Fall ab, Herz vom Baum der Zeit,
fallt, ihr Blätter, aus den erkalteten Ästen,
die einst die Sonne umarmt',
fallt, wie Tränen fallen aus dem geweiteten Aug!

Fliegt noch die Locke taglang im Wind
um des Landgotts gebräunte Stirn,
unter dem Hemd preßt die Faust
schon die klaffende Wunde.

Drum sei hart, wenn der zarte Rücken der Wolken
sich dir einmal noch beugt,
nimm es für nichts, wenn der Hymettos die Waben
noch einmal dir füllt.

Denn wenig gilt dem Landmann ein Halm in der Dürre,
wenig ein Sommer vor unserem großen Geschlecht.

Und was bezeugt schon dein Herz?
Zwischen gestern und morgen schwingt es,
lautlos und fremd,
und was es schlägt,
ist schon sein Fall aus der Zeit.

Manuella Blackburn
The Fortune Teller, The Crane aus *Origami* 2007
These two miniatures are taken from a larger work entitled Origami focusing on the Japanese paper-folding art form. This is my second work to make use of a compositional tool developed from Denis Smalley's spectromorphological language, this time specifically focusing on different types of motions. Experimentation with this vocabulary informed the creation of directional, reciprocal and cyclic motions that the origami structural shapes initially inspired.
Origami is the art of economy – a few simple folds can suggests an animal or shape and with slight modifications an entirely different creation can appear.

1. *The Fortune Teller* (otherwise known as paper-foldy-thing) is a representation of regularity. Its final symmetrical form is flexible – stretching outward and collapsing inwards, while its function, as a method of concealing and revealing hidden fortunes, is presented as a game of chance.

2. *The Crane* is a representation of good fortune. It is an agile bird with a fleeting presence and swift movements.

Ioannis Kalantzis *Parastaseis A B C D* 2003-2006

Four electroacoustic miniatures.
This work is based on the realisation of sound organisms, which lead across their movements and their unusual logic to a particular type of energy. The sound material is based in recordings that I made from strings, clarinet, bassoon and guitar. The elaboration of the material includes digital sound processing techniques as filtering, cross synthesis, transposition and sound synthesis. I also used a combination of panning and amplitude changes in order to give to the sound a rather energetic movement. The principal transformation of the sound is made by Audio Sculpt, Metasynth and MaxMSP.

Annette Vande Gorne *Yawar Fiesta (opera), 2. part*
Opera on a booklet from Werner Lambersy
Women' Chorus, Act II : Combattimento 10:00 2007
Soprano: Françoise Vanhecke
Contraltos: Fadila figuidi, Annette Vande Gorne
Space format: 7.1
Can space, written and performed live, lead to expression and dramatization? Music? Opera? Even if the project does not disclose any singer on stage, which will be human and silent, a great deal of the sound material relates to human voice. It is not an electroacoustic work (with its abstract materials and writings reseaches), but the dramatization of a text by notably placing objects and spatial shapes.
The booklet, written on an incantatory and ritual mode of antic tragedies, reflects dual fights we are facing, civilisations fights symbolised by those of eagles and bulls in Andean village. Those for instance of appealling desire and rough strength. In the bourgeoisy ladies chorus of Act II, dramatic singing and lyrism came step by step to the surface, with an unstoppable clearness. Result therefrom an hybrid writing, that combines spaces, energies, morphologes and harmonic colors.

Produced at Studio Métamorphoses d'Orphée, Musiques & Recherches, Ohain (Belgique).
Creation of Chorus II: Bruxelles, théâtre Le Marni, festival « L'Espace du Son », 17 octobre 2007

Chœur II
Taureau, nous quitterons pour toi
l'habit de lumière,
car la génisse de l'aube, nous l'avons entendu gémir
sous les coups de ta croupe,
et le lait de tes reins,
nous l'avons vu fumer dans les vallées brumeuses du matin ;
ton piétinement de sabot,
nous en avons nourri nos âmes ;
nous rêvons aux verges nocturnes du mystère ;
voici que nous brosserons, du gant de crin de nos cris, ton courage.
Cela, pour que luise l'obscur
et soit peigné et présentable le poème
à la robe revêche et rude de ta vitalité.

Contents

Session 1: Artificial Intelligence, Machine Learning, Genetic Algorithms

Session 2: Spatialisation Technologies

Session 3: Perception of Space

Session 4: Performance and Interaction

Session 5: Space in Sound. Aesthetical Aspects

Session 6: Space in Sound. Musicological Aspects

Session 7: Algorithms and Environments for Music Composition

Session 8: Studio Reports and Environments

Session 9: Feature Extraction and Classification

Session 10: Sound Synthesis

An A-Life Approach to Machine Learning of Musical Worldviews for Improvisation Systems

Marcelo Gimenes and Eduardo R. Miranda
Interdisciplinary Centre for Computer Music Research
University of Plymouth, UK
(marcelo.gimenes, eduardo.miranda)@plymouth.ac.uk

Abstract — **In this paper we introduce Interactive Musical Environments (iMe), an interactive intelligent music system based on software agents that is capable of learning how to generate music autonomously and in real-time. iMe belongs to a new paradigm of interactive musical systems that we call "ontomemetical musical systems" for which a series of conditions are proposed.**

I.Introduction

Tools and techniques associated with Artificial Life (A-Life), a discipline that studies natural living systems by simulating their biological occurrence on computers, are an interesting paradigm that deals with extremely complex phenomena. Actually, the attempt to mimic biological events on computers is proving to be a viable route for a better theoretical understanding of living organisms [1].

We have adopted an A-Life approach to intelligent systems design in order to develop a system called iMe (Interactive Music Environment) whereby autonomous software agents perceive and are influenced by the music they hear and produce. Whereas most A-Life approaches to implementing computer music systems are chiefly based on algorithms inspired by biological development and evolution (for example, Genetic Algorithms [2]), iMe is based on cultural development (for example, Imitation Games [3, 4]).

Central to iMe are the notions of musical style and musical worldview. Style, according to a famous definition proposed by Meyer, is "a replication of patterning, whether in human behaviour or in the artefacts produced by human behaviour, that results from a series of choices made within some set of constraints" [5]. Patterning implies the sensitive perception of the world and its categorisation into forms and classes of forms through cognitive activity, "the mental action or process of acquiring knowledge and understanding through thought, experience and the senses" (Oxford Dictionary).

Worldview, according to Park [6], is "the collective interpretation of and response to the natural and cultural environments in which a group of people lives. Their assumptions about those environments and the values derived from those assumptions." Through their worldview people are connected to the world, absorbing and exercising influence, communicating and interacting with it. Hence, a musical worldview is a two-way route that connects individuals with their musical environment.

In our research we want to tackle the issue of how different musical influences can lead to particular musical worldviews. We therefore developed a computer system that simulates environments where software agents interact among themselves as well as with external agents, such as other systems and humans. iMe's general characteristics were inspired in the real world: agents perform musical tasks for which they possess perceptive and cognitive abilities. Generally speaking, agents perceive and are influenced by music. This influence is transmitted to other agents as long as they generate new music that is then perceived by other agents, and so forth.

iMe enables the design and/or observation of chains of musical influence similarly to what happens with human musical apprenticeship. The system addresses the perceptive and cognitive issues involved in musical influence. It is precisely the description of a certain number of musical elements and the balance between them (differences of relative importance) that define a musical style or, as we prefer to call it, a musical worldview: the musical aesthetics of an individual or of a group of like-minded individuals (both, artificial and natural).

iMe is referred to as an ontomemetic computer music system. In Philosophy of Science, ontogenesis refers to the sequence of events involved in the development of an individual organism from its birth to its death. However, our research is concerned with the development of cultural organisms rather than biological organisms. We therefore coined the term "ontomemetic" by replacing the affix "genetic" by the term "memetic". The notion of "meme" was suggested by Dawkins [7] as the cultural equivalent of gene in Biology. Musical ontomemesis therefore refers to the sequence of events involved in the development of the musicality of an individual.

An ontomemetic musical system should foster interaction between entities and, at the same time, allow for the observation of how different paths of development can lead to different musical worldviews. Modelling perception and cognition abilities plays an important role in our system, as we believe that the way in which music is perceived and organized in our memory has direct connections with the music we make and appreciate. The more we get exposed to certain types of elements, the more these elements get meaningful representations in our memory. The result of this exposure and interaction is that our memory is constantly changing, with new elements being added and old elements being forgotten.

Despite the existence of excellent systems that can learn to simulate musical styles [8] or interact with human performers in real-time ([9-11]), none of them address the problem from the ontomemetic point of view, i.e.:

- to model perceptive and cognitive abilities in artificial entities based on their human correlatives
- to foster interaction between these entities as to nurture the emergence of new musical worldviews

- to model interactivity as ways through which reciprocal actions or influences are established
- to provide mechanisms to objectively compare different paths and worldviews in order to assess their impact in the evolution of a musical style.

An ontomemetic musical system should be able to develop its own style. This means that we should not rely on a fixed set of rules that restrain the musical experience to particular styles. Rather, we should create mechanisms through which musical style could eventually emerge from scratch.

In iMe, software entities (or agents) are programmed with identical abilities. Nevertheless, different modes of interactions give rise to different worldviews. The developmental path, that is the order in which the events involved in the development of a worldview takes place, plays a crucial role here. Paths are preserved in order to be reviewed and compared with other developmental paths and worldviews. A fundamental requisite of an ontomemetic system is to provide mechanisms to objectively compare different paths and worldviews in order to assess the impact that different developmental paths might have had in the evolution of a style. This is not trivial to implement.

A. *Improvisation*

Before we introduce the details of iMe, a short discussion about musical improvisation will help to better contextualise our system. Not surprisingly, improvised music seems to be a preferred field when it comes to the application of interactivity, and many systems have been implemented focusing on controllers and sound synthesis systems designed to be operated during performance. The interest in exploring this area, under the point of view of an ontomemetic musical system relies on the fact that, because of the intrinsic characteristics of improvisation, it is intimately connected with the ways human learning operates. However, not many systems produced for music improvisation to date are able to learn.

According to a traditional definition, musical improvisation is the spontaneous creative process of making music while it is being performed. It is like speaking or having a conversation as opposed to reciting a written text.

As it encompasses musical performance, it is natural to observe that improvisation has a direct connection with performance related issues such as instrument design and technique. Considering the universe of musical elements played by improvisers, it is known that certain musical ideas are more adapted to be played with polyphonic (e.g., piano, guitar) as opposed to monophonic instruments (e.g., saxophone, flute) or with keyboards as opposed to wind instruments, and so forth.

Since instrument design and technique affect the easiness or difficulty of performing certain musical ideas, we deduce that different musical elements must affect the cognition of different players in different ways.

The technical or "performance part" of a musical improvisation is, at the same time, passionate and extremely complex but, although we acknowledge the importance of its role in defining one's musical worldview, our research (and this paper) is focused primarily on how: (i) music is perceived by the sensory organs, (ii) represented in memory and (iii) the resulting cognitive processes relevant to musical creation in general (and more specifically, to improvisation) conveys the emergence and development of musical worldviews.

Regarding specifically the creative issue, it is important to remember that improvisation, at least in its most generalised form, follows a protocol that consists of developing musical ideas "on top" of pre-existing schemes. In general, these include a musical theme that comprises, among other elements, melody and harmonic structure. Therefore, in this particular case, which happens to be the most common, one does not need to create specific strategies for each individual improvisational session but rather follow the generally accepted protocol.

Despite of the fact that this may give the impression to be limiting the system, preventing the use of more complex compositional strategies, one of the major interests of research into music improvisation relies on the fact that once a musical idea has been played, one cannot erase it. Therefore, each individual idea is an "imposition" in itself that requires completion that leads to other ideas, which themselves require completion, and so on. Newly played elements complete and re-signify previous ones in such ways that the improviser's musical worldview is revealed. In this continuous process two concurrent and different plans play inter-dependent roles: a pathway (the "lead sheet") to which the generated ideas have to adapt and the "flow of musical ideas" that is particular to each individual at each given moment and that imply (once more) their musical worldview.

The general concepts introduced so far are all an integral part of iMe and will be further clarified as we introduce the system.

II.The iMe System

iMe was conceived to be a platform in which software agents perform music related tasks that convey musical influence and emerge their particular styles. Tasks such as read, listen, perform, compose and improvise have already been implemented; a number of others are planned for the future. In a multi-agent environment one can design different developmental paths by controlling how and when different agents interact; a hypothetical example is shown in Fig. 1.

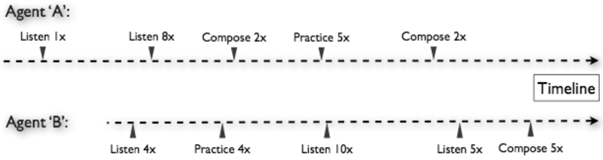


Fig. 1. The developmental paths of two agents.

In the previous figure we see the representation of a hypothetical timeline during which two agents (Agent 'A' and Agent 'B') perform a number of tasks. Initially, Agent 'A' would listen to one piece of music previously present in the environment. After that, Agent 'B' would listen to 4 pieces of music and so forth until one of them, Agent 'A' would start to compose its own pieces. From this moment Agent 'B' would listen to the pieces composed by Agent 'A' until Agent 'B' itself would start to compose and then Agent 'A' would interact with Agent 'B's music as well.

In general, software agents should normally act autonomously and decide if and when to interact. Nevertheless, in the current implementation of iMe we decided to constrain their skills in order to have a better control over the development of their musical styles:

agents can choose which music they interact with but not how many times or when they interact.

When agents perform composition or improvisation tasks, new pieces are delivered to the environment and can be used for further interactions. On the other hand, by performing tasks such as read or listen to music, agents only receive influence.

Interaction can be established not only amongst the agents themselves, but also between agents and human musicians. The main outcome of these interactions is the emergence and development of the agents' musical styles as well as the musical style of the environment as a whole.

The current implementation of iMe's perceptive algorithms was specially designed to take into account a genre of music texture (homophonic) in which one voice (the melody) is distinguishable from the accompanying harmony. In the case of the piano for instance, the player would be using the left hand to play a series of chords while the right hand would be playing the melodic line. iMe addresses this genre of music but also accepts music that could be considered a subset of it; e.g., a series of chords, a single melody or any combination of the two. Any music that fits into these categories should generate an optimal response by the system. However, we are also experimenting with other types of polyphonic music with a view on widening the scope of the system.

In a very basic scenario, simulations can be designed by simply specifying:

- A number of agents
- A number of tasks for each agent
- Some initial music material for the interactions

iMe generates a series of consecutive numbers that correspond to an abstract time control (cycle). Once the system is started, each cycle number is sent to the agents, which then execute the tasks that were scheduled to that particular cycle.

As a general rule, when an agent chooses a piece of music to read (in the form of a MIDI file) or is connected to another agent to listen to its music, it receives a data stream which is initially decomposed into several feature streams, and then segmented as described in the next section.

A. *System's Perception and Memory*

iMe's perception and memory mechanisms are greatly inspired by the work of Snyder [12] on musical memories. According to Snyder, "the organisation of memory and the limits of our ability to remember have a profound effect on how we perceive patterns of events and boundaries in time. Memory influences how we decide when groups of events end and other groups of events begin, and how these events are related. It also allows us to comprehend time sequences of events in their totality, and to have expectations about what will happen next. Thus, in music that has communication as its goal, the structure of the music must take into consideration the structure of memory - even if we want to work against that structure".

iMe's agents initially "hear" music and subsequently use a number of filters to extract independent but interconnected streams of data, such as melodic direction, melodic inter-onset intervals, and so on. This results in a feature data stream that is used for the purposes of segmentation, storage (memory) and style definition (Fig. 2).

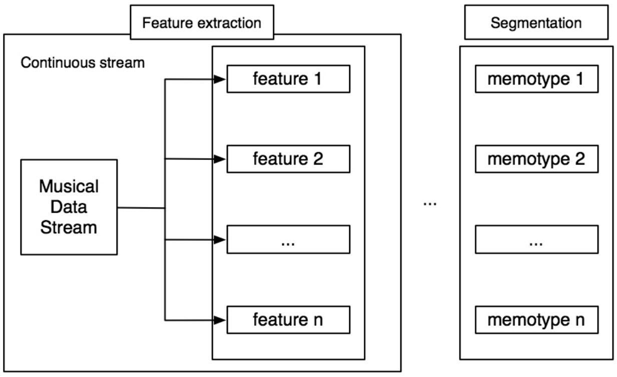
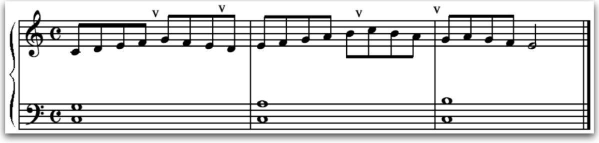


Fig. 2. Feature extraction and segmentation.

To date we have implemented ten filters, which extract information from melodic (direction, leap, inter-onset interval, duration and intensity) and non-melodic notes (vertical number of notes, note intervals from the melody, inter-onset interval, duration and intensity). As it might be expected, the higher the number of filters, the more accurate is the representation of the music. In order to help clarify these concepts, in Fig. 3 we present a simple example and give the corresponding feature data streams that would have been extracted by an agent, using the ten filters:



	1	2	3	4	5	6	7	8	9	10	11	...
a)	0	1	1	1	1	-1	-1	-1	1	1	1	...
b)	0	2	2	1	2	2	1	2	2	1	2	...
c)	120	120	120	120	120	120	120	120	120	120	120	...
d)	120	120	120	120	120	120	120	120	120	120	120	...
e)	6	6	6	6	6	6	6	6	6	6	6	...
f)	2	0	0	0	0	0	0	0	2	0	0	...
g)	5, 7	-2	-2	-2	-2	-2	-2	-2	7, 9	-2	-2	...
h)	120	-2	-2	-2	-2	-2	-2	-2	120	-2	-2	...
i)	960	-2	-2	-2	-2	-2	-2	-2	960	-2	-2	...
j)	6	-2	-2	-2	-2	-2	-2	-2	6	-2	-2	...

Fig. 3. Feature streams, where a) melody direction, b) melody leap, c) melody interonset interval, d) melody duration, e) melody intensity, f) non melody number of notes, g) non melody note intervals from melody, h) non melody interonset interval, i) non melody duration, j) non melody intensity.

Number -2 represents the absence of data in a particular stream. Melody direction can value -1, 0 and 1, meaning descending, lack of and ascending movement, respectively. Melody leaps and intervals are shown in half steps. In streams that hold time information (interonset intervals and duration) the value 240 (time resolution) is assigned to quarter notes. Intensity is represented by the MIDI range (0 to 127); in Fig. 3 this was simplified by dividing this value by ten.

After the extraction of the feature data stream, the next step is the segmentation of the music. A fair amount of research has been conducted on this subject by a number of scholars. In general, the issue of music segmentation remains unsolved to a great extent due to its complexity. One of the paradigms that substantiate segmentation systems has been settled by Gestalt psychologists who argued that perception is driven from the whole to the parts by the application of concepts that involve simplicity and uniformity in organising perceptual information [13].

Proximity, closure, similarity and good continuation are some of these concepts.

Fig. 4 shows a possible segment from piece by J. S. Bach (First Invention for Two Voices) according to Gestalt theory. In this case the same time length separates all except for the first and the last notes, which are disconnected from the previous and the following notes by rests. This implies the application of similarity and proximity rules.

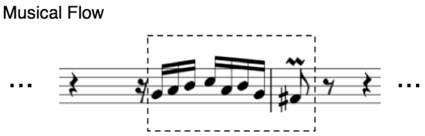


Fig. 4. An example of a music segment.

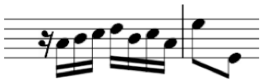
In the example discussed below we decided to build the segmentation algorithm on top of only one of the principles that guide group organization: the occurrence of surprise. As the agents perceive the continuous musical stream by the various expert sensors (filters), wherever there is a break in the continuity of the behaviour of one (or a combination of some) of the feature streams, this is an indication of positions for a possible segmentation. The whole musical stream is segmented at these positions. If discontinuities happen in more than one feature at the same time, this indicates the existence of different levels of structural organization within the musical piece; this conflict must be resolved (this will be clarified later).

In the example of Fig. 3, we shall only consider the melody direction stream ('a' of Fig. 3). Hence, every time the direction of the melody is about to change, a new grouping starts. These places are indicated on the musical score shown in Fig. 3 with the symbol 'v'.

To designate these segmented musical structures we adopted the expression "musical meme" or simply "meme", a term that has been introduced by Dawkins [7] to describe basic units of cultural transmission in the same way that genes, in biology, are units of genetic information. "Examples of memes are tunes, catch-phrases, clothes fashions, ways of making pots or of building arches. Just as genes propagate themselves in the gene pool by leaping from body to body via sperm and eggs, so memes propagate in the meme pool by leaping from brain to brain via a process which, in a broad sense, can be called imitation." [7].

The idea of employing this concept is attractive because it covers both the concept of structural elements and processes of cultural development, which fits well with the purpose of our research.

A meme is generally defined as a short musical structure, but it is difficult to ascertain what is the minimal acceptable size for a meme. In iMe, memes are generally small structures in the time dimension and they can have any number of simultaneous notes. Fig. 5 shows a meme (from the same piece of the segment shown in Fig. 4) and its memotype representation following the application of three filters: melodic direction, leap and duration:



Mel. direction:	0	1	1	1	-1	1	-1	1	-1
Mel. leap:	0	2	2	1	3	2	4	7	12
Mel. duration:	0	60	60	60	60	60	60	120	120

Fig. 5. Meme and corresponding memotype representation.

Since the memes were previously separated into streams of data, they can be represented as a group of memotypes, each corresponding to a particular musical feature. A meme is therefore represented by 'n' memotypes, in which 'n' is the number of streams of data representing musical features. In any meme the number of elements of all the memotypes is the same and corresponds to the number of vertical structures. By "vertical structure" we mean all music elements that happen at the same time.

B. Memory

The execution of any of the musical tasks requires the perception and segmentation of the musical flow and the adaptation of the memory. As a result, the agents need to store this information in their memory by comparing it with the elements that were previously perceived. This is a continuous process that constantly changes the state of the memory of the agents.

In iMe, the memory of the agents comprises a Short Term Memory (STM) and a Long Term Memory (LTM). The STM consists of the last x memes (x is defined "a priori" by the user) that were most recently brought to the agent's attention, representing the focus of their "awareness".

A much more complex structure, the LTM is a series of specialized "Feature Tables" (FTs), a place designed to store all the memotypes according to their categories. FTs are formed by "Feature Lines" (FLs) that keep a record of the memotypes, the dates of when the interactions took place (date of first contact - dfc, date of last contact - dlc), the number of contacts (noc), weight (w) and "connection pointers" (cp). In Fig. 6 we present the excerpt of a hypothetical FT (for melody leaps) in which there are 11 FLs. The information between brackets in this Fig. corresponds to the memotype and the numbers after the colon correspond to the connection pointers. This representation will be clarified by the examples given later.

Feature n. 2 (melody leaps):	
Line 0:	[0 0]: 0 0 0 0 0 0 0 0 0 0
Line 1:	[2 2 0 1 0 1 2 5 0]: 1
Line 2:	[1 0 0 3 2 2 0]: 2 20 10 10
Line 3:	[1 0 0 0 1 2 2 4]: 3
Line 4:	[2 0 2 0 4 1 3 0]: 4
Line 5:	[0 3 2 7 0 2 0 4]: 5 8 10
Line 6:	[3 0 2 0 3 2 4]: 6 5 3
Line 7:	[1 0 1 2 2 0 3]: 7 3
Line 8:	[2 0 2 0 2 0 0]: 8 31 8
Line 9:	[2 0]: 47 4 9 9 4 9 9
Line 10:	[5 0 8 2 1 2]: 10

Fig. 6. A Feature Table excerpt.

1) Adaptation

Adaptation is generally accepted as one of the cornerstones of evolutionary theories, Biology and indeed A-Life systems. With respect to cultural evolution, however, the notion of adaptation still seem to generate heated debates amongst memetic theory scholars. Cox [14] asserts that the "memetic hypothesis" is based on the concept that the understanding that someone has on sounds comes from the comparison with the sounds already produced by this person. The process of comparison would involve tacit imitation, or memetic participation that is based on the previous personal experience on the production of the sound.

According to Jan [15] "the evolution of music occurs because of the differential selection and replication of mutant memes within idioms and dialects. Slowly and incrementally, these mutations alter the memetic configuration of the dialect they constitute. Whilst gradualistic, this process eventually leads to fundamental changes in the profile of the dialect and, ultimately, to seismic shifts in the overarching principles of musical organization, the rules, propagated within several dialects."

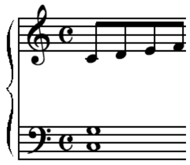
iMe defines that every time agents interact with a piece of music their musical knowledge changes according to the similarities and/or differences that exist between this piece and their own musical "knowledge". At any given time, each memotype for each one of the FTs in an agent's memory is assigned with a weight that represents their relative importance in comparison with the corresponding memotypes in the other memes.

The adaptation mechanism is fairly simple: the weight is increased when a memotype is perceived by an agent. The more an agent listens to a memotype, the more its weight is increased. Conversely, if a memotype is not listened to for some time, its weight is decreased; in other words, the agent begins to forget it.

The forgetting mechanism - an innovation if compared to other systems, such as the ones cited earlier - is central to the idea of an ontomemetic musical system and is responsible for much of the ever-changing dynamics of the weights of memotypes. In addition to this mechanism, we have implemented a "coefficient of permeability" (values between 0 and 1) that modulates the calculation of the memotype weights. This coefficient is defined by a group of other variables (attentiveness, character and emotiveness), the motivation being that some tasks entail more or less transformation to the agent's memory depending on the required level of attentiveness (e.g., a reading task requires less attention than an improvisation task). On the other hand, attributes such as character and emotiveness can also influence the level of "permeability" of the memory.

When a new meme is received by the memory, if the memotype is not present in the corresponding FT, a new FL is created and added to the corresponding FT. The same applies to all the FTs in the LTM. The other information in the FLs (dates, weight and pointers) is then (re)calculated. This process is exemplified below.

Let us start a hypothetical run in which the memory of an agent is completely empty. As the agent starts perceiving the musical flow (Fig. 3), the agent's "sensory organs" (feature filters) generate a parallel stream of musical features, according to the mechanism described earlier. The first meme (Fig. 7) then arrives at the agent's memory and, as a result, the memory is adapted (Fig. 8).



Feature stream:
mdi: 0, 1, 1, 1
mle: 0, 2, 2, 1
mii: 120, 120, 120, 120
mdu: 120, 120, 120, 120

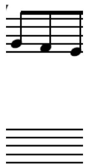
Fig. 7. Meme 1, where mdi is melody direction, mle is melody leap, mii is melody interonset interval and mdu is melody duration.

In order to keep the example simple, we are only showing the representation of four selected features: melody direction (FT1), leap (FT2), interonset interval (FT3) and duration (FT4). Fig. 8 shows the memotypes in each of the Feature Tables. Notice that the connection pointers (cp) of FTs 2 to 4 actually point to the index (i) of the memotype of FT1. The initial weight (w) was set to 1.0 for all of the memotypes and the information date (dfc, dlc) refers to the cycle in which this task is performed during the simulation; in this case, the first task.

i	Memotype	dfc	dlc	noc	w	cp
Melody direction:						
1	0, 1, 1, 1	1	1	1	1.0	
Melody leap:						
1	0, 2, 2, 1	1	1	1	1.0	1
Melody interonset interval:						
1	120, 120, 120, 120	1	1	1	1.0	1
Melody duration:						
1	120, 120, 120, 120	1	1	1	1.0	1

Fig. 8. Agent's memory after adaptation to meme 1.

Then comes the next meme (Fig. 9), as follows:



Feature stream:
mdi: 1, -1, -1
mle: 2, 2, 1
mii: 120, 120, 120
mdu: 120, 120, 120

Fig. 9. Meme 2.

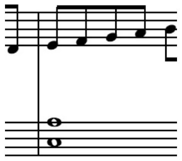
And the memory is adapted accordingly (Fig. 10):

i	Memotype	Dfc	dlc	noc	w	cp
Melody direction:						
1	0, 1, 1, 1	1	1	1	1.0	2
2	1, -1, -1	1	1	1	1.0	
Melody leap:						
1	0, 2, 2, 1	1	1	1	1.0	1
2	2, 2, 1	1	1	1	1.0	2
Melody interonset interval:						
1	120, 120, 120, 120	1	1	1	1.0	1
2	120, 120, 120	1	1	1	1.0	2
Melody duration:						
1	120, 120, 120, 120	1	1	1	1.0	1
2	120, 120, 120	1	1	1	1.0	2

Fig. 10. Agent's memory after adaptation to meme 2.

Here all the new memotypes are different from the previous ones and stored in separate FLs in the corresponding FTs. Now the memotype of index 1 in FT1 points (cp) to the index 2. Differently from the other FTs, this information represents the fact that memotype of index 2 comes after the memotype of index 1. This shows how iMe keeps track of the sequence of memes to which the agents are exposed. The cp of the other FTs still point to the index in FT1 that connect the elements of the meme to which the memory is being adapted. The weights of the new memes are set to 1.0 as previously.

The same process is repeated with the arrival of meme 3 (Figs. 11 and 12) and meme 4 (Figs. 13 and 14).



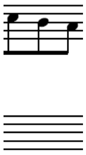
Feature stream:
mdi: -1, 1, 1, 1, 1, 1
mle: 2, 2, 1, 2, 2, 2
mii: 120, 120, 120, 120, 120, 120
mdu: 120, 120, 120, 120, 120, 120

Fig. 11. Meme 3.

i	Memotype	dfc	dlc	Noc	W	Cp
Melody direction:						
1	0, 1, 1, 1	1	1	1	1.0	2
2	1, -1, -1	1	1	1	1.0	3
3	-1, 1, 1, 1, 1, 1	1	1	1	1.0	
Melody leap:						
1	0, 2, 2, 1	1	1	1	1.0	1

2	2, 2, 1	1	1	1	1.0	2
3	2, 2, 1, 2, 2, 2	1	1	1	1.0	3
	Melody interonset interval:					
1	120, 120, 120, 120	1	1	1	1.0	1
2	120, 120, 120	1	1	1	1.0	2
3	120, 120, 120, 120, 120, 120	1	1	1	1.0	3
	Melody duration:					
1	120, 120, 120, 120	1	1	1	1.0	1
2	120, 120, 120	1	1	1	1.0	2
3	120, 120, 120, 120, 120, 120	1	1	1	1.0	3

Fig. 12. Agent's memory after adaptation to Meme 3.



Feature stream:
mdi: 1, -1, -1
mle: 1, 1, 2
mii: 120, 120, 120
mdu: 120, 120, 120

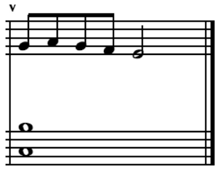
Fig. 13. Meme 4.

The novelty here is that the memotypes for melody direction, interonset interval and duration had already been stored in the memory. Only the melody leap has new information and, as a result a new FL was added to FT2 and not to the other FTs. The weights of the repeated memotypes were increased by '0.1', which means that the relative weight of this information increased if compared to the other memotypes. We can say thereafter that the weights ultimately represent the relative importance of all the memotypes in relation to each other. The memotype weight is increased by a constant factor (e,g, f = 0.1) every time it is received and decreases by another factor if, at the end of the cycle, it is not "perceived". The later case will not happen in this example because we are considering that the run is being executed entirely in one single cycle.

i	Memotype	dfc	dlc	noc	W	Cp
	Melody direction:					
1	0, 1, 1, 1	1	1	1	1.0	2
2	1, -1, -1	1	1	2	1.1	3
3	-1, 1, 1, 1, 1, 1	1	1	1	1.0	2
	Melody leap:					
1	0, 2, 2, 1	1	1	1	1.0	1
2	2, 2, 1	1	1	1	1.0	2
3	2, 2, 1, 2, 2, 2	1	1	1	1.0	3
4	1, 1, 2	1	1	1	1.0	2
	Melody interonset interval:					
1	120, 120, 120, 120	1	1	1	1.0	1
2	120, 120, 120	1	1	2	1.1	2, 2
3	120, 120, 120, 120, 120, 120	1	1	1	1.0	3
	Melody duration:					
1	120, 120, 120, 120	1	1	1	1.0	1
2	120, 120, 120	1	1	2	1.1	2, 2
3	120, 120, 120, 120, 120, 120	1	1	1	1.0	3

Fig. 14. Agent's memory after adaptation to meme 4.

Finally, the memory receives the last meme (Fig. 15) and is adapted accordingly (Figs. 15 and16).



Feature stream:
mdi: -1, 1, -1, -1, -1
mle: 2, 2, 2, 2, 1
mii: 120, 120, 120, 120, 120
mdu: 120, 120, 120, 120, 480

Fig. 15. Meme 5.

i	memotype	dfc	dlc	noc	w	cp
	Melody direction:					
1	0, 1, 1, 1	1	1	1	1.0	2
2	1, -1, -1	1	1	2	1.1	3, 4
3	-1, 1, 1, 1, 1, 1	1	1	1	1.0	2
4	-1, 1, -1, -1, -1	1	1	1	1.0	
	Melody leap:					
1	0, 2, 2, 1	1	1	1	1.0	1
2	2, 2, 1	1	1	1	1.0	2
3	2, 2, 1, 2, 2, 2	1	1	1	1.0	3
4	1, 1, 2	1	1	1	1.0	2
5	2, 2, 2, 2, 1	1	1	1	1.0	4
	Melody interonset interval:					
1	120, 120, 120, 120	1	1	1	1.0	1
2	120, 120, 120	1	1	2	1.1	2, 2
3	120, 120, 120, 120, 120, 120	1	1	1	1.0	3
4	120, 120, 120, 120, 120	1	1	1	1.0	4
	Melody duration:					
1	120, 120, 120, 120	1	1	1	1.0	1
2	120, 120, 120	1	1	2	1.1	2, 2
3	120, 120, 120, 120, 120, 120	1	1	1	1.0	3
4	120, 120, 120, 120, 480	1	1	1	1.0	4

Fig. 16. Agent's memory after adaptation to meme 5.

C. Generative Processes

Gabora [16] explains that, in the same way that information patterns evolve through biological processes, mental representation - or memes - evolves through the adaptive exploration and transformation of an informational space through variation, selection and transmission. Our minds perform tasks on its replication through an aptitude landscape that reflects internal movements and a worldview that is continuously being updated through the renovation of memes.

In iMe agents are also able to compose through processes of re-synthesis of the different memes from their worldview. Obviously, the selection of the memes that will be used in a new composition implies that the musical worldview of this agent is also re-adapted by reinforcing the weights of the memes that are chosen.

In addition to compositions (non real-time), agents also execute two types of real-time generative tasks: solo and collective improvisations. The algorithm is described below.

1) Solo improvisations

During solo improvisations, only one agent play at a time, following the steps below

a) Step 1: Generate a new meme according to the current "meme generation mode"

The very first memotype of a new piece of music is chosen from the first Feature Table (FT1), which guides de generation of the whole sequence of memes, in a Markov-like chain. Let us assume that the user configured FT1 to represent melody direction. Hence, this memotype could be, hypothetically [0, 1, 1, -1], where 0 represents "repeat the previous note", 1 represents upward motion and -1 represents downward motion. Once the memotype from FT1 is chosen (based on the distribution of probability of the weights of the memotypes in that table), the algorithm looks at the other memotypes at the other FTs to which the memotype at FT1 points at and chooses a memotype for each FT of the LTM according to the distribution of probability of the weights at each FT. At this point we would end up with a new meme (a series of n memotypes, where n = number of FTs in the LTM).

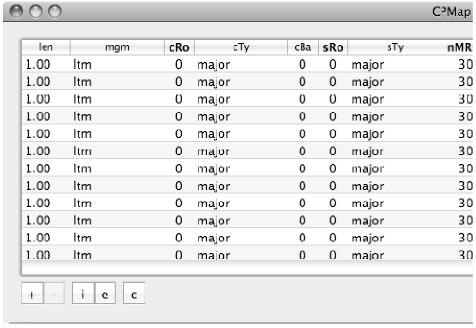
The algorithm of the previous paragraph describes one of the generation modes that we have implemented: the "LTM generation mode". There are other modes. For instance, there is the "STM generation mode", where agents choose from the memes stored in their Short Term Memory. Every time a new meme is generated, the agent checks the Compositional and Performance Map (explanation below) to see which generation mode is applicable at any given time.

b) Step 2: Adapt the memory with the newly generated meme

Once the new meme is generated, the memory is immediately adapted to reflect this choice, according to the criteria explained in the previous section.

c) Step 3: Adapt the meme to the Compositional and Performance Map (CPM)

The new meme is then adapted according to criteria foreseen at the CPM. The CPM (Fig. 17), iMe's equivalent to a “lead sheet”, possesses instructions regarding a number of parameters that address both aspects of the improvisation: the generation of new musical ideas and the performance of these ideas. Examples of the former are: the meme generation mode, transformations to the meme, local scales and chords, note ranges for right and left hand. Examples of the latter are: ratio of loudness between melodic and non-melodic notes, shifts for note onset, loudness and duration both for melodic and non-melodic notes. Instructions regarding the performance only affect the sound that is generated by the audio output of the system and is not stored with the composition.



len	mgm	cRo	cTy	cBa	sRo	sTy	nMR
1.00	ltm	0	major	0	0	major	30
1.00	ltm	0	major	0	0	major	30
1.00	ltm	0	major	0	0	major	30
1.00	ltm	0	major	0	0	major	30
1.00	ltm	0	major	0	0	major	30
1.00	ltm	0	major	0	0	major	30
1.00	ltm	0	major	0	0	major	30
1.00	ltm	0	major	0	0	major	30
1.00	ltm	0	major	0	0	major	30
1.00	ltm	0	major	0	0	major	30
1.00	ltm	0	major	0	0	major	30
1.00	ltm	0	major	0	0	major	30

Fig. 17. A CPM excerpt.

The instructions (or “constraints”) contained in the CPM are distributed on a timeline. The agent checks the constraints that are applicable at the “compositional pointer”, a variable that controls the position of the composition on the timeline, and acts accordingly.

d) Step 4: Generate notes and play the meme (if in real time mode)

Until this moment, the memes are not real notes but only meta-representations described by the memotypes (melody direction, melody leap, etc.). Given the previously generated notes and the CPM, the “actual notes” of the meme must be calculated and sent to a playing buffer.

e) Step 5: Store the meme in the composition

An array with the information of the sequence of the memes is kept with the composition for future reference and tracking of the origin of each meme. There is another generation mode, the “MemeArray generation mode”, where an agent can retrieve any previously generated meme and choose it again during the composition.

f) Step 6: Repeat previous steps until the end of the CPM

The agent continuously plays the notes of the playing buffer. When the number of notes in this buffer is equal to or less than 'x' (parameter configured by the user), the algorithm goes back to step 1 above and a new meme is generated until the whole CPM is completed.

2) Collective improvisations

The steps for collective improvisations are very similar to the steps for solo improvisations, except for the fact that the agents play along with a human being. We have implemented this task as two separate sub-tasks (a listening sub-task and a solo improvisation sub-task) running in separate threads. Memes are generated as in a solo improvisation and the agents' memory is equally affected by the memes they choose as well as by the memes that they listen from the musical data originated by the external improviser. Both agent and external improviser follow the same CPM.

At the end of the improvisation (solo or interactive), the composition is stored in the system in order to be used in further runs of the system.

III. Conclusions and Further Work

In this paper we introduced Interactive Musical Environments (iMe) for the investigation of the emergence and evolution of musical styles in environments inhabited by artificial agents, under the perspective of human perception and cognition. This system belongs to a new paradigm of interactive musical systems that we refer to as “ontomemetical musical systems” for which we propose a series of prerequisites and applications.

As seen from some of the experiments that we have presented, we understand that iMe has the potential to be extremely helpful in areas such as the musicological investigation of musical styles and influences. Besides the study of the development of musical styles in artificial worlds, we are also conducting experiments with human subjects in order to assess iMe's effectiveness to evaluate musical influences in inter-human interaction. The study of creativity and interactive music in artificial and real worlds could also benefit with a number of iMe's features, which we are currently evaluating as well.

The memory of an agent is complex and dynamic, comprising of all memotypes, their weights and connection pointers. The execution of musical tasks affects the memory state in proportion to the appearance of different memes and memotypes. A particular musical ontomemesis can thereafter be objectively associated with the development of any agent's “musicality”.

Bearing in mind that iMe can be regarded as a tool for the investigation of musical ontomemesis as much as a tool for different sorts of musicological analyses, a series of different simulation designs could be described.

Future improvements to the system will include the introduction of algorithms that would allow iMe to become a self-sustained artificial musical environment such as criteria to control the birth and demise of agents and the automatic definition of their general characteristics such as attentiveness, character, emotiveness, etc. Agents should also possess the ability to decide when and what tasks to perform, besides being able to develop their own Compositional and Performance Maps.

Acknowledgment

The authors would like to thank the funding support from the Brazilian Government's Fundacao Coordenacao de Aperfeicoamento de Pessoal de Nivel Superior (CAPES).

REFERENCES

1. Miranda, E.R., *The artificial life route to the origins of music.* Scientia, 1999. **10**(1): p. 5-33.
2. Biles, J.A. *GenJam: A Genetic Algorithm for Generating Jazz Solos.* in *International Computer Music Conference*. 1994.
3. Miranda, E.R., *Emergent Sound Repertoires in Virtual Societies.* Computer Music Journal, 2002. **26**(2): p. 77-90.
4. Miranda, E.R., *At the Crossroads of Evolutionary Computation and Music: Self-Programming Synthesizers, Swarm Orchestras and the Origins of Melody.* Evolutionary Computation, 2004. **12**(2): p. 137-158.
5. Meyer, L.B., *Style and Music: Theory, History, and Ideology*. 1989, Philadelphia: University of Pennsylvania Press.
6. Park, M.A., *Introducing Anthropology: An Integrated Approach*. 2002: McGraw-Hill Companies.
7. Dawkins, R., *The Selfish Gene*. 1989, Oxford: Oxford University Press.
8. Cope, D., *Computers and Musical Style*. 1991, Oxford: Oxford University Press.
9. Rowe, R., *Interactive Music Systems: Machine Listening and Composing*. 1993: MIT Press.
10. Pachet, F., *Musical Interaction with Style.* Journal of New Music Research, 2003. **32**(3): p. 333-341.
11. Assayag, G., et al. *Omax Brothers: a Dynamic Topology of Agents for Improvization Learning*. in *Workshop on Audio and Music Computing for Multimedia, ACM Multimedia*. 2006. Santa Barbara.
12. Snyder, B., *Music and Memory: An Introduction*. 2000, Cambridge, MA: MIT Press.
13. Eysenck, M.W. and M.T. Keane, *Cognitive Psychology: A Student's Handbook*. 2005: Psychology Press.
14. Cox, A., *The mimetic hypothesis and embodied musical meaning.* MusicæScientiæ, 2001. **2**: p. 195–212.
15. Jan, S., *Replicating sonorities: towards a memetics of music.* Journal of Memetics - Evolutionary Models of Information Transmission, 2000. **4**.
16. Gabora, L., *The Origin and Evolution of Culture and Creativity.* Journal of Memetics, 1997.

Breeding Rhythms with Artificial Life

Joao M. Martins and Eduardo R. Miranda
Interdisciplinary Centre for Computer Music Research
University of Plymouth, UK
{joao.martins, eduardo.miranda}@plymouth.ac.uk

Abstract — **We are interested in developing intelligent systems for music composition. In this paper we focus on our research into generative rhythms. We have adopted an Artificial Life (A-Life) approach to intelligent systems design in order to develop generative algorithms inspired by the notion of music as social phenomena that emerge from the overall behaviour of interacting autonomous software agents. Whereas most A-Life approaches to implementing computer music systems are chiefly based on algorithms inspired by biological evolution (for example, Genetic Algorithms [2]), this work is based on cultural development (for example, Imitation Games [12, 13]). We are developing a number of such "cultural" algorithms, one of which is introduced in this paper: the *popularity algorithm*. We also are developing a number of analysis methods to study the behaviour of the agents. In our experiments with the popularity algorithm we observed the emergence of coherent repertoires of rhythms across the agents in the society.**

I.INTRODUCTION

The A-Life approach to music is a promising new development for composers. It provides an innovative and natural means for generating musical ideas from a specifiable set of primitive components and processes reflecting the compositional process of generating a variety of ideas by brainstorming followed by selecting the most promising ones for further iterated refinement [8]. We are interested in implementing systems for composition using A-Life-based models of cultural transmission; for example, models of the development and maintenance of musical styles within particular cultural contexts, and their reorganization and adaptation in response to cultural exchange.

Existing A-Life-based systems for musical composition normally employ a Genetic Algorithm (GA) to produce musical melodies, rhythms, and so on. In these systems, music parameters are represented as "genotypes" and GA operators are applied on these representations to produce music according to given fitness criteria. Because of the highly symbolic nature of Western music notation, music parameters are suitable for GA-based processing and a number of musicians have used such systems to compose music.

Although we acknowledge that there have been a few rather successful stories [2], we believe that additional A-Life-based methods need to be developed [11, 12]. The work presented in this paper contributes to these developments by looking into the design of algorithms that consider music as a cultural phenomenon whereby social pressure plays an important role in the development of music. A plausible method to embed social dynamics in such algorithms is to design them within the framework of interacting autonomous software agents.

We are developing a multi-agent system for composition of rhythms where the user will be able to extract information about the behaviour of the agents and the evolving rhythms in many different ways, providing composers the means to explore the outcomes systematically. An in-depth discussion on the architecture of the whole system and how it will be used artistically to compose pieces of music falls beyond the scope of this paper. Rather, this paper will focus on one of the A-Life algorithms that we have developed for the system - the *popularity algorithm* - and the information that one can extract about its behaviour, and the analyses of the behaviours.

By way of related research, we cite the work by de Boer [3] on modelling the emergence of vowel systems by means of imitations games and Kirby's work on evolution of language [9]. Also, Miranda [13] has developed a model of the emergence of intonation systems using imitation games. Basically an imitation game consists of one agent picking a random sound from its repertoire and the other agent trying to imitate it. Then, a feedback is given about the success of the imitation. On the basis of this feedback, the agents update their memories.

II.THE AGENTS

The agents are identical to each other and the number of agents in a group may vary. The agents move in a virtual 2D space and they normally interact in pairs. Essentially, the agents interact by playing rhythmic sequences to each other, with the objective of developing repertoires of rhythms collectively. At each round, each of the agents in a pair plays one of two different roles: the *player* and the *listener*. The agents may perform operations on the rhythms that they play to each other, depending on the iteration algorithm at hand and on the status of the emerging repertoire. The agents are provided with a memory to store the emerging rhythms and other associated information.

The fundamental characteristic of human beings is that we are able to perceive, and more importantly, to produce an isochronous pulse [6]. Moreover, humans show a preference for rhythms composed of integer ratios of the basic isochronous pulse [5]. Therefore, we represent rhythms as interonset intervals in terms of small integer ratios of an isochronous pulse (Fig. 1).

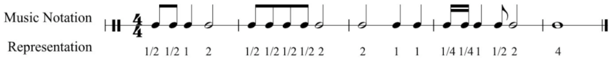


Fig. 1. Standard music notation of a rhythmic sequence and its corresponding interonset representation.

A. Transformations of Rhythms

At the core of the mechanism by which the agents develop rhythmic sequences is a set of basic transformation operations. These operations enable the agents to generate new rhythmic sequences and change the rhythmic sequences that they learn as the result of the interactions with other agents. The transformation operations are as follows:

- Divide a rhythmic figure by two (Fig. 2a)
- Merge two rhythmic figures (Fig. 3b)
- Add one element to the sequence (Fig. 2c)
- Remove one element from the sequence (Fig. 2d)

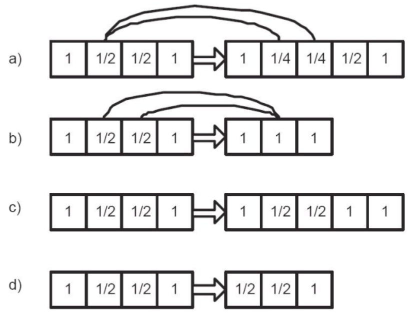


Fig. 2. Examples of rhythmic transformations.

The definition of these transformations were inspired by the dynamical systems approach to study human bimanual coordination [7] and is based on the notion that two coupled oscillators will converge to stability points at frequencies related by integer ratios [1]. We have defined other transformations that divide a figure into three, five, and other prime numbers, but the impact of these additional transformations on the system is beyond the scope of this paper. Addition and removal transformations were introduced to increase diversity in the pool of rhythms and to produce rhythms of different lengths.

B. Measuring Similarity of Rhythms

The agents are programmed with the ability to measure the degree of similarity of two rhythmic sequences. This measurement is used when they need to assess the similarity of the rhythms they play to each other. Also, this algorithm is used to measure the similarity between repertoires of rhythms from different agents.

In a previous paper [10] we introduced a method to measure the degree of similarity between two sequences of symbols by comparing various subsequences at various levels. The result is a vector, referred to as the *Similarity Coefficients Vector* (SCV), which contains the interim results of the comparisons between subsequences. For the present work, we devised a version of the SCV method to deal with rhythmic sequences.

Let us define the block distance between two sequences containing the same number of elements as follows:

$$d(\mathbf{v}, \mathbf{w}) = \sum_{i=1}^{n} |v_i - w_i|$$

where **v** and **w** are the two sequences (vectors) that are being compared, and v_i and w_i are the individual components of these vectors. After obtaining the resulting evaluation of the block distances on a given level (length of a subsequence), we can write a matrix for the *k*-level, corresponding to the comparison of all the subsequences with length *k* between the two main sequences:

$$\mathbf{D}^{(k)} = \begin{bmatrix} d(\mathbf{v}_1^{(k)}, \mathbf{w}_1^{(k)}) & \dots & d(\mathbf{v}_1^{(k)}, \mathbf{w}_{(m-k+1)}^{(k)}) \\ d(\mathbf{v}_2^{(k)}, \mathbf{w}_1^{(k)}) & \dots & d(\mathbf{v}_2^{(k)}, \mathbf{w}_{(m-k+1)}^{(k)}) \\ \vdots & \vdots & \vdots \\ d(\mathbf{v}_{(n-k+1)}^{(k)}, \mathbf{w}_1^{(k)}) & \dots & d(\mathbf{v}_{(n-k+1)}^{(k)}, \mathbf{w}_{(m-k+1)}^{(k)}) \end{bmatrix}$$

where *d* are the distances $d(\mathbf{v}, \mathbf{w})$ between all the subsequences $\mathbf{v}^{(k)}$ of **v** and all the subsequences $\mathbf{w}^{(k)}$ of **w**. Next, let us define the *k*-level *Similarity Coefficient* as follows:

$$c^{(k)}(\mathbf{v}, \mathbf{w}) = \frac{z(k)}{(n-k+1)(m-k+1)}$$

where *z(k)* is the number of zeros in the matrix $\mathbf{D}^{(k)}$. Roughly speaking, the similarity coefficient measures the sparsity of the matrix $\mathbf{D}^{(k)}$. The higher the coefficient *c(k)*, the higher is the similarity between the subsequences of level *k*. Next, we can collect all the *k*-levels coefficients in a vector referred to as *Similarity Coefficient Vector* (SCV). This is defined as follows:

$$\mathbf{C} = \left[c^{(1)}, c^{(2)}, \dots, c^{(min(m,n))}\right]$$

Fig. 3 shows an example of building a 3-level Distance Matrix and its respective SCV is *SCV* = [0.4167 0.1333 0.1250 0]. From SCV we can obtain a scalar value in order to establish a comparative analysis between larger sets of rhythms, such as the repertoires of two agents. We can take the rightmost nonzero value from the SCV, which corresponds to the higher level where two matching sequences can be found. We can either take a weighted sum of the SCV values or the average of all values, as follows:

$$SCV_{av} = \frac{1}{min(m,n)} \sum_{k=1}^{min(m,n)} SCV(k)$$

where *SCV* (*k*) are the coefficients of similarity for each of the *k* levels. The next step is to compare the repertoire of the agents in order to observe the development of relationships amongst the agents in a group of agents; for instance, to observe if the agents form distinct sub-groupings. The similarity of the repertoire of rhythms amongst the agents in a group is computed by creating a matrix of SCV_{av} values of the repertoires of all pairs of agents. Matrices with the columns and rows corresponding to the number of rhythms in the memory of each agent reveal the similarity between their repertoires (Fig. 4).

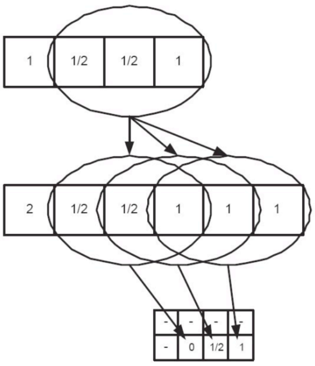


Fig. 3. Example of building a 3-level Distances Matrix.

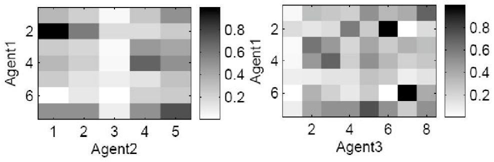


Fig. 4. Examples of similarity matrices between the repertoires of 3 agents: agent 1 vs. agent 2 and agent 1 vs. agent 3. The darker the colour the more similar the rhythms are.

By collapsing both the rows and the columns of the matrices, and taking the maximum values for each of them and an averaged sum, we obtain the scalar of similarity between repertoires, as follows:

$$SimRep_{k,l} = \frac{1}{nR_{Ak}+nR_{Al}}\left[\sum_{i=1}^{nR_{Ak}} max(SCV_{av})_{rows} + \sum_{j=1}^{nR_{Al}} max(SCV_{av})_{cols}\right]$$

where the first term corresponds to the sum of the maximum values of the SCV_{av}, for every row, and the second term is the correspondent for every column; nR_{Ak} and nR_{Al} are the number of rhythms in the repertoire of the compared agents.

Finally, the development of repertoires of rhythms of a group of agents as a whole can be observed by conducting a hierarchical cluster analysis of all distance measures between the agents (*DistRep*). This cluster analysis produces a dendrogram using a linkage method based on an unweighted average distance, also known as group average in which the distance between two clusters *A* and *B*, D_{AB}, is given by the following equation:

$$D_{AB} = \frac{1}{N_A.N_B}\sum_i d_i$$

where N_A and N_B are the number of elements in *A* and *B*, and d_i are pairwise distances between the elements of clusters *A* and *B*. The hierarchical cluster analysis produces a dendrogram of the type shown in Fig. 8. Such dendrogram is drawn through an iterative process until all the individuals or clusters are linked.

C. Measuring the Complexity of Rhythms

The complexity of a rhythmic sequence is measured as follows:

$$Complexity = \frac{nF + \sum_{i=1}^{nF} n_i}{\sum_{i=1}^{nF} T_i}$$

where *nF* is the number of rhythmic figures contained in the sequence, n_i is value of the numerator of a rhythmic figure, and T_i is the relative length of a rhythmic figure, considering that each rhythmic figure is a fraction of the pulse. This is a computationally cost effective method to measure the complexity of a rhythmic sequence.

It is important to bear in mind that our implementation ensures that there are no reducible fractions included in the sequence, meaning that there is always a single numerical representation for a given rhythm. Fig. 5 shows an example of a graph plotting the complexity of a sequence of relative interonset intervals [1, 1] as it is transformed thirty times recurrently, using the transformation operations mentioned earlier.

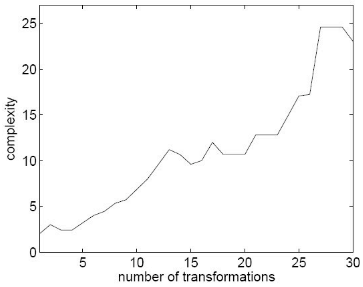


Fig. 5. Example where complexity increases with the number of transformations.

III. The Popularity Algorithm and Experiments

Popularity is a numerical parameter that each agent attributes to a rhythm in its repertoire. This parameter is modified both by the (agent-)listener and by the (agent-)player during the interactions. If the listener recognises a rhythm (that is, if it holds the “perceived” rhythm in its repertoire), then it will increase the popularity index of this rhythm and will give a positive feedback to the player. A positive feedback is an acknowledgment signal, which will prompt the player to increase the popularity index of the rhythm in question in its repertoire. Conversely, if the listener does not recognize the rhythm, then it will add it to its repertoire and will give a negative feedback to the player. This negative feedback will cause the player to decrease the popularity index of this rhythm.

Furthermore, there is a memory loss mechanism whereby after each interaction all the rhythms have their popularity index decreased by 0.05. This accounts for a natural drop in the popularity index due to ageing. The core of the popularity algorithm works as follows:

```
Agent Player:
    P1. Plays a rhythm and increase the counter for the
    number of times that this rhythm has been used.
Agent Listener:
    L1. Search for the heard rhythm in its repertoire
    L2. If the rhythm is found, then give a positive feedback to
    the agent player and increase the counter for the
    popularity of the rhythm in its repertoire
    L3. If the rhythm is not found, then add this rhythm to the
    repertoire and give as negative feedback to the agent
    player
Agent Player:
    P2. Receive the feedback from agent listener
    P3. If feedback is positive, then increase the counter for
    the popularity of the rhythm in its repertoire
    P4. If feedback is negative, then decrease the counter for
    the popularity algorithm in its repertoire
    P5. If the minimum popularity threshold for this rhythm
    has been reached, then remove this rhythm from its
    repertoire
    P6. If the transformation threshold for this rhythm has
    been reached, then transform this rhythm
```

As for the analyses, firstly we analyse the development of the size and the complexity of the repertoire of individual agents. Then, we analyse the values of the corresponding individual measures from the agents, as well as similarity between agents and how they are clustered in terms of the rhythms they share. Finally, we measure the lifetime of the rhythms, the amount of rhythmic sequences that the society develops and the degree to which the agents share similar rhythms. We trace the lifetime of a rhythm by counting the number of agents that hold the sequence in their memories during the interactions. Fig. 6 shows 3 examples of analyses.

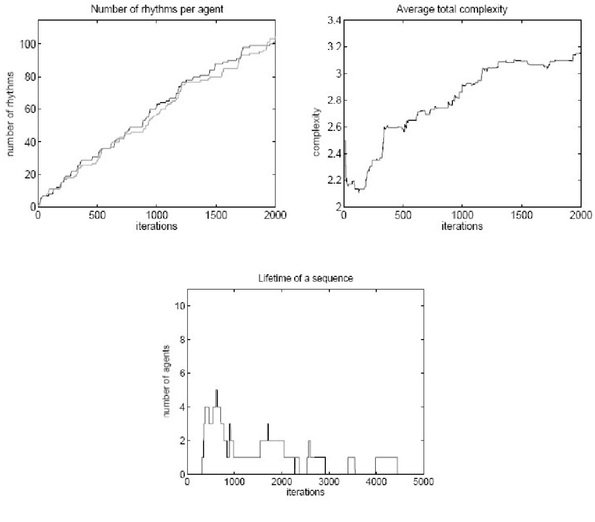


Fig. 6. Examples of amalyses: development of the size of the repertoire for different agents (top left), complexity of the rhythms of the society (top right) and number of agents sharing a particular rhythm (bottom).

The experiments were run for 5000 iterations each for a number of times, with the objective of observing the behaviour of the agents, the society and the evolving rhythms, under different conditions. We have run experiments with societies of 3, 10 and 50 agents. For some of the experiments we have limited the lifetime of the agents to 1000 iterations; when an agent dies, another is born. Sometimes the algorithms take into account the movement of the agents in the 2D space, which may or may not influence the nature of the iterations. Fig. 7 shows the results after 5000 iterations of the popularity algorithm with 10 agents (without population renewal).

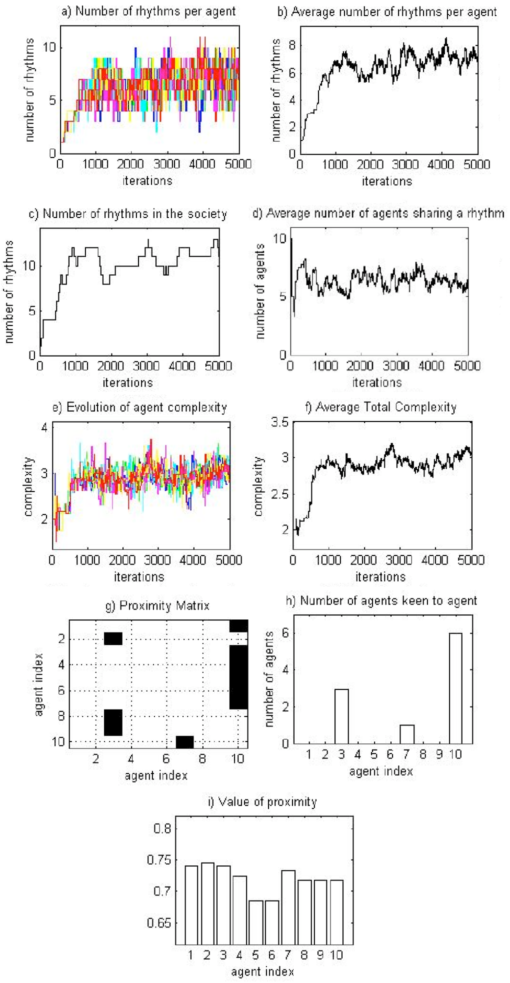


Fig. 7. Results from a typical run of the popularity algorithm with 10 agents.

Fig. 7a displays the development of the repertoire of individual agents and Fig. 7b displays the corresponding average across all agents. Here the repertoires of the agents grow steadily up to approximately 1000 iterations and subsequently oscillates around a stable point. Fig. 7c displays the development of the repertoire of the whole society being a direct consequence of the lifetime of each rhythm. The average number agents sharing a rhythm (Fig. 7d) is calculated by summing the instant number of agents sharing a rhythm for all rhythms, and dividing the result by the number of rhythms currently present in the society (Fig. 7c). This graph (Fig. 7d) provides the means to assess the global behaviour of the society; for instance, if it develops coherently in terms of popularity of existing rhythms. Fig. 7e represents the development of complexity of the individual agents and Fig. 7f gives the corresponding average. Initially, the size and complexity of the repertoire of individual agents are very close to the average, but this trend is replaced quickly by repertoires of different sizes amongst the agents.

The last three graphs show the degree of similarity between the repertoires of the agents according to the similarity measure defined earlier. Fig. 7g displays information about the identity of the agent with whom each agent relates most; i.e., has the highest similarity value. The graph in Fig. 7h shows the agents that are regarded by others as being most similar to them. In this case, it shows that agent 3 has three agents with similar repertoires, and agent 10 is the one that concentrates the highest number of keen agents, having six agents considering its repertoire to be more similar to theirs.

Hierarchical cluster analysis is conducted in order to observe groupings of agents according to the similarity of their repertoires. Fig. 8 shows the dendrogram containing elements of three societies of 10 agents each: Society 1 comprises agents 1 to 10, Society 2 comprises agents 11 to 20 and Society 3 the remaining 21 to 30. By comparing the three societies we can observe 3 clearly independent clusters, which were developed separately in three separate runs with the same set of parameters. In addition to the previous observations, this suggests that the repertoires that emerged from the popularity algorithm display diversity, are stable in terms of size, and are coherent within their respective societies. We can also observe differences in the clusters within a given society.

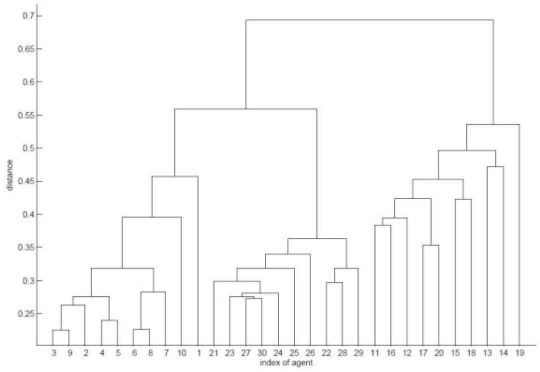


Fig. 8. Dendrogram resulting from the hierarchical cluster analysis conducted in the context of the popularity algorithm containing three independent societies with 10 agents each.

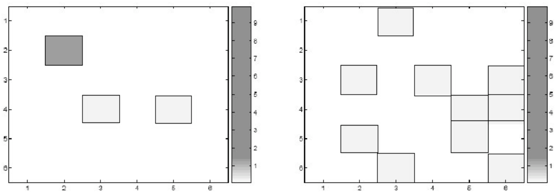


Fig. 9. World visualisation of two steps of the iterative process. Clustering takes place (figure on the left) followed by scattering at a later stage (figure on the right). A cluster is indicated by a darker colour.

By letting the agents move in their environment, we also investigated whether the interaction rules could influence the movement of the agents and whether this process would influence the development of their repertoires. In this case, if a listening agent “recognises” the rhythm played by the other agent, then it will follow the player agent in the space in the next iteration. Fig. 9 shows periodic clustering of one or more groups of agents that move together and keep interacting until the cluster is scattered due to an unsuccessful interaction.

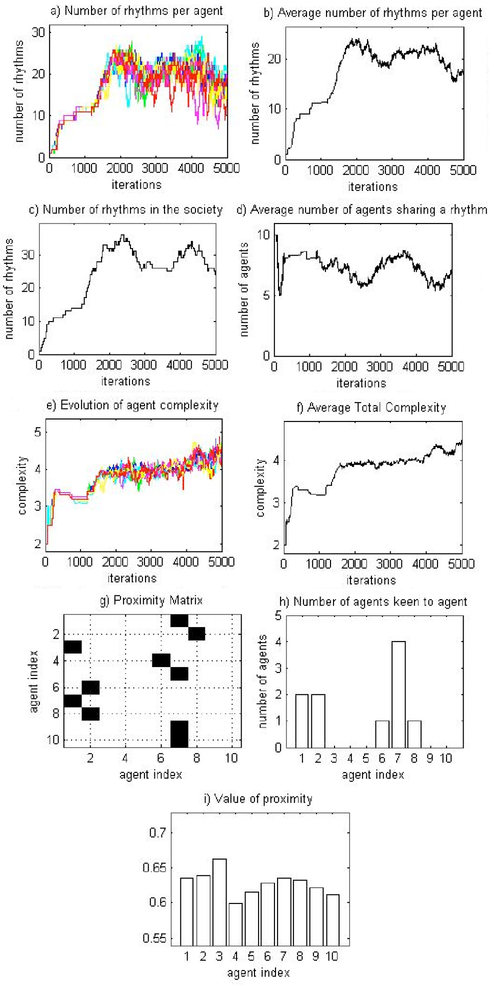


Fig. 10. Results from a typical run of the popularity algorithm taking into account the movement of the agents as an influencing factor in the evolution of the repertoire.

In Fig. 10, we can observe two behaviours that are typical of the popularity algorithm with movement taken into account. The first being that there are many more rhythms affecting the interactions than in the case without movement; this is due to the fact that every time a positive feedback occurs, two or more agents will form a group. This increases the number of interactions and consequently the number of rhythms in their repertoires. The second being that there is an initial overshoot of the size of the repertoire before reaching a level of stability. This is possibly caused by the initial clustering of agents when individual repertoires grow consistently among very closely related agents.

IV. CONCLUDING DISCUSSSION

We are developing novel A-Life-based generative music algorithms with a view on producing an intelligent system for the composition of rhythms. Most current approaches to using A-Life in software for generating music entail the application of a GA. We propose that a strictly GA-based approach to generate music is questionable because they were not designed to address musical problems in the first place, but to solve engineering and searching problems. The act of composing music seldom involves an automated selective procedure towards an ideal outcome based on a set of definite fitness criteria.

As a way forward, we suggest that A-Life-based systems for generating music should employ algorithms that consider music as a cultural phenomenon whereby social pressure plays an important role in the development of musical conventions. To this end, we are developing a number of algorithms, one of which was introduced in this paper: the *popularity algorithm*. In addition, we developed a number of methods to monitor the behaviour of the algorithms.

In all runs of the popularity algorithm we observed the emergence of coherent repertoires across the agents in the society. Clustering of agents according to their repertoires could also be observed on various occasions.

Whereas the size of the repertoire is controlled by a popularity parameter in the algorithm, it tends to grow constantly in the other algorithms that we have implemented. We also observed that a small subset of agents tend to concentrate the preference of most of the population. This trend tended to appear in many runs with different settings.

ACKNOWLEDGMENT

This project was developed thanks to a grant from The Leverhulme Trust.

REFERENCES

[1] Beek, P. J., Peper, C. E., and Daffertshofer, A. (2000). “Timekeepers versus nonlinear oscillators: how the approaches differ”. In P. Desain and L. Windsor, Eds., *Rhythm Perception and Production*, pp. 9-34. London: Swets and Zeitlinger (now Taylor & Francis).

[2] Biles, J. A. (1994). “Genjam: A genetic algorithm for generating jazz solos”. *Proceedings of the International Computer Music Conference*, Aarhus(Denmark). San Francisco, USA: International Computer Music Association.

[3] de Boer, B. (1999). *Self-Organisation in Vowel Systems*. PhD thesis, Vrije Universiteit Brussel.

[4] Doornbusch, P. (2005). *The Music of the CSIRAC: Australia’s First Computer Music*. Victoria, Australia: Common Ground.

[5] Drake, C. and Bertrand, D. (2001). “The Quest for Universals in Temporal Processing in Music”. *Annals of the New York Academy of Sciences*, 930(1):17-27.

[6] Handel, S. (1989). *Listening: An Introduction to the Perception of Auditory Events*. Cambridge, USA: The MIT Press.

[7] Kelso, J. A. (1984). “Phase transitions and critical behavior in human bimanual coordination”. *American Journal of Physiology – Regulatory, Integrative and Comparative Physiolpgy*, 246(6):1000-1004.

[8] Kim, K.-J. and Cho, S.-B. (2006). “A comprehensive overview of the applications of artificial life”. *Artificial Life*, 12(1):153-182.

[9] Kirby, S. (2002). “Natural language from artificial life”. *Artificial Life*, 8(2):185-215.

[10] Martins, J. M., Gimenes, M., Manzolli, J., and Maia Jr., A. (2005). “Similarity measures for rhythmic sequences”. *Proceedings of the 10th Brazilian Symposium on Computer Music (SBCM)*, Belo Horizonte (Brazil).

[11] Miranda, E. and Biles, A., Eds. (2007). *Evolutionary Computer Music*. London, UK: Springer-Verlag.

[12] Miranda, E. R. (2004). "At the Crossroads of Evolutionary Computation and Music: Self-Programming Synthesizers, Swarm Orchestras and the Origins of Melody", *Evolutionary Computation* 12(2):137-158.

[13] Miranda, E. R. (2002). “Mimetic development of intonation”, *Proceedings of the 2nd International Conference on Music and Artificial Intelligence (ICMAI 2002)*. Springer Verlag - Lecture Notes on Artificial Intelligence.

Melody Characterization by a Genetic Fuzzy System

Pedro J. Ponce de León*, David Rizo*, Rafael Ramirez†, José M. Iñesta*
*Departamento de Lenguajes y Sistemas Informáticos, Universidad de Alicante, Spain
†Music Technology Group, Universitat Pompeu-Fabra, Barcelona, Spain

***Abstract*—We present preliminary work on automatic human-readable melody characterization. In order to obtain such a characterization, we (1) extract a set of statistical descriptors from the tracks in a dataset of MIDI files, (2) apply a rule induction algorithm to obtain a set of (crisp) classification rules for melody track identification, and (3) automatically transform the crisp rules into fuzzy rules by applying a genetic algorithm to generate the membership functions for the rule attributes. Some results are presented and discussed.**

I. Introduction

Melody is a somewhat elusive musical term that often refers to a central part of a music piece that catches most of the listener's attention, and which the rest of music parts are subordinated to. This is one of many definitions that can be found in many places, particularly music theory manuals. However, these are all formal but subjective definitions given by humans. The goal in this work is to automatically obtain an objective and human friendly characterization of what it is considered to be a melody.

The identification of a melody track is relevant for a number of applications like melody matching [1], motif extraction from score databases, or extracting melodic ringtones from MIDI files. In this work we approach the problem of automatically building a model that characterizes melody tracks. Such a model is tested in experiments on finding a melody track in a MIDI file. The melody model is a set of human-readable fuzzy rules automatically induced from a corpora of MIDI files by using statistical properties of the musical content.

To our best knowledge, the automatic description of a melody has not been tackled as a main objective in the literature. The most similar problem to the automatic melody definition is that of finding a melody line from a polyphonic source. This problem has been approached mainly for three different objectives and with different understandings of what a melody is. The first objective is the extraction of the melody from a polyphonic audio source. For this task it is important to describe the melody in order to leave out those notes that are not candidates to belong to the melody line[2]. In the second objective, a melody line (mainly monophonic) must be extracted from a symbolic polyphonic source where no notion of *track* is used [3]. The last objective is to select one track containing the melody from a list of input tracks of symbolic polyphonic music (e.g. MIDI). Ghias et al. [1] built a system to process MIDI files extracting a sort of melodic line using simple heuristics. Tang et al. [4] presented a work where the aim was to propose candidate melody tracks, given a MIDI file. They take decisions based on single features derived from informal assumptions about what a melody track may be. Madsen and Widmer [5] try to solve the problem by the use of several combination of the entropies of different melody properties like pitch classes, intervals, and IOI.

A. What's a melody?

Before focusing on the machine learning methodology to extract automatically the characterization of a *melody*, the musical concept of melody needs to be reviewed.

Melody is a concept that has been given many definitions, all of them complementary. The variability of the descriptions can give an idea on the difficulty of the task to extract a description automatically.

From the music theory point of view, Ernst Toch [6] defines it as "a succession of different pitch sounds brighten up by the rhythm". He also writes "a melody is a sound sequence with different pitches, in opposition to its simultaneous audition that constitutes what is named as chord". He distinguishes also the term "melody" from the term "theme".

A music dictionary [7] defines melody as: "a combination of a pitch series and a rhythm having a clearly defined shape".

An informal survey was carried out where the subjects were asked to answer the question *What is a melody?*. Both musicians and non-musicians took part in the survey. The following list is a *compendium* of shared melody traits found in answers gathered on that survey:

- (finite) succession of notes
- *cantabile* pitch range
- monophonic
- lead part
- identifies/characterices the piece, song
- unity
- diversity
- contains repeating patterns
- often linked to text
- done by humans
- understandable, memorizable by humans

The music theory literature lacks the same amount of works about melody than can be found about counterpoint, harmony, or "form" [8]. Besides, the concept of melody is dependant on the genre or the cultural convention. The most interesting studies about melody have appeared in recent years, mainly influenced by new emerging models like generative grammars [9], artificial intelligence [10], and Gestalt and cognitive psychology [11]. All these works place effort on understand the melody in order to generate it automatically.

The types of tracks and descriptions of *melody* versus *accompaniment* is posed in [8]. The author distinguishes:

- *compound* melodies where there is only a melodic line where some notes are principal, and others tend to accompany, being this case the most frequent in unaccompanied string music.
- *self-accompanying* melodies, where some pitches pertain both to the thematic idea and to the harmonic (or rhythmic) support
- *submerged* melodies consigned to inner voices
- *roving* melodies, in which the theme migrates from part to part
- *distributed* melodies, in which the defining notes are divided between parts and the prototype cannot be isolated in a single part.

From the audio processing community, several definitions can be found about what a melody is. Maybe, the most general definition is that of Kim et at. [12]: “melody is an auditory object that emerges from a series of transformations along the six dimensions: pitch, tempo, timbre, loundness, spatial location, and reverberan environment".

Gómez et al. [13] gave a list of mid and low-level features to describe melodies:

- Melodic attributes derived from numerical analysis of pitch information: number of notes, tessitura, interval distribution, melodic profile, melodic density.
- Melodic attributes derived from musical analysis of the pitch data: key information, scale type information, cadence information.
- Melodic attributes derived from a structural analysis: motive analysis, repetitions, patterns location, phrase segmentation.

Another attempt to describe a melody can be found in [14]. In that book, Temperley proposes a model of melody perception based on three principles:

- Melodies tend to remain within a narrow pitch range.
- Note-to-note intervals within a melody tend to be small.
- Notes tend to conform to a key profile (a distribution) that depends on the key.

All these different properties a melody should have can be a reference to compare the automatic results.

The rest of the paper is organized as follows: first, the methodology used in this work is presented. Second, the experimentation framework is outlined. Next, results on several datasets for both crisp and fuzzy rule systems are discussed and compared to related work results. Finally, conclusions and further work are presented.

II. Methodology

The goal of this work is to obtain an human-readable characterization of MIDI tracks containing melody lines, against other kind of tracks. A fuzzy rule system has been chosen as the technique to obtain such a characterization. These fuzzy models should achieve good performance in discriminating melody tracks when compared to other non-fuzzy or non-rule based crisp models.

The methodology applied to obtain such fuzzy models is sketched as follows: first, MIDI tracks are described by a set of statistical features on several properties of the track content. This is presented in section II-A. Next section briefly describes different rule extraction methods used to obtain crisp rule systems that characterize melody tracks. Finally, these rule systems are then converted to fuzzy rule systems applying a fuzzyfication process to the input domain. This is discussed in section II-C.

A. MIDI track content description

MIDI track content is described by a collection of statistics on several properties of musical note streams, such as pitch, pitch interval or note duration, as well as track properties such as number of notes in the track, track duration, polyphony rate or occupation rate. As a result, MIDI tracks are represented by vectors $v \in \mathbb{R}^{34}$ of statistical values. This representation has been used to characterize melody tracks in previous works [15], [16].

This set of statistical descriptors is presented in Table I. The first column indicates the category being analyzed, and the second one shows the kind of statistics describing properties from that category. The third column indicates the range of the descriptor[1].

Four features were designed to describe the track as a whole and fifteen to describe particular aspects of its content. For the latter descriptors, both normalized and non-normalized versions have been computed. Only non-normalized ones are displayed in table I. Normalized descriptors are defined in $[0,1]$ and computed using the formula

$$(v_i - min)/(max - min)$$

where v_i is the descriptor value to be normalized corresponding to the i-th track, and min and max are, respectively, the minimum and maximum values for this descriptor for all the tracks of the target midifile. This allows to represent these properties proportionally

[1] $[x..y]$ denotes integer domains and $[x, y]$ denotes real domains.

TABLE I

MIDI TRACK DESCRIPTORS

Category	Descriptors	Domain
Track info.	Normalized duration Number of notes Occupation rate Polyphony rate	[0,1] [0 .. +∞[[0,1] [0,1]
Pitch	Highest Lowest Mean Standard deviation	[0 .. 127] [0 .. 127] [0,127] [0,+∞[
Pitch intervals	Number of distinct intv. Largest Smallest Mean Mode Standard deviation	[0 .. 127] [0 .. 127] [0 .. 127] [0,127] [0 .. 127] [0,+∞[
Note durations	Longest Shortest Mean Standard deviation	[0,+∞[[0,+∞[[0,+∞[[0,+∞[
Syncopation	No. of syncopated notes	[0 .. +∞[
Class	IsMelody	{true, false}

to other tracks in the same file, using non-dimensional values. This way, a total number of $4 + 15 \times 2 = 34$ descriptors were initially computed for each track.

The track information descriptors are normalized duration (using the same scheme as above), number of notes, occupation rate (proportion of the track length occupied by notes), and the polyphony rate (the ratio between the number of ticks in the track where two or more notes are active simultaneously and the track duration in ticks).

Pitch descriptors are measured using MIDI pitch values. The maximum possible MIDI pitch is 127 (pitch G_8) and the minimum is 0 (pitch C_{-2}).

The interval descriptors summarize information about the difference in pitch between consecutive notes. Absolute pitch interval values are computed.

Finally, note duration descriptors are computed in terms of beats, so they are independent from the MIDI file resolution. *Syncopations* are notes that start at some place between beats (usually in the middle) and extend across them.

B. A rule system for melody characterization

In this work, a rule system obtained using the RIPPER algorithm [17] is used as the basis to induce a fuzzy rule system. Briefly, the RIPPER constructs a rule set RS by considering each class, from the less prevalent one to the more frequent one. It builds RS until the description length (DL) of the rule set and examples is 64 bits greater than the smallest DL met so far, or there are no positive examples, or the error rate >= 50%. Rules are constructed by greedily adding antecedents to the rule until the rule is perfect (i.e. 100value of each attribute and selects the condition with highest information gain (for details see [17]). We applied the RIPPER algorithm and obtained a rule system from the SMALL dataset (see section III), so it is called the RIPPER-SMALL rule system. Table II shows the rules in this system. Note that only 13 out of 34 initial statistical descriptors have been selected by the algorithm to characterize melody tracks. Figures about this rule system performance are presented in section V.

TABLE II

RIPPER-SMALL (CRISP) RULES.

Name	Rule
R1	if (AvgPitch >= 65.0) and (TrackOccupationRate >= 0.51) and (AvgAbsInterval <= 3.64) and (TrackNumNotes >= 253) then IsMelody=true
R2	if (AvgPitch >= 62.6) and (TrackOccupationRate >= 0.42) and (TrackPolyphonyRate <= 0.21) and (NormalizedDistinctIntervals >= 1) then IsMelody=true
R3	if (AvgPitch >= 65.4) and (TrackNumNotes >= 284) and (ShortestNormalizedDuration <= 0.001) and (ShortestDuration >= 0.02) and (NormalizedDistinctIntervals >= 1) then IsMelody=true
R4	if (AvgAbsInterval <= 2.72) and (TrackSyncopation >= 16) and (AvgPitch >= 60.5) and (TrackOccupationRate >= 0.42) and (StdDeviationPitch <= 5.0) then IsMelody=true
R5	if (AvgAbsInterval <= 3.87) and (TrackSyncopation >= 24) and (LowestNormalizedPitch >= 0.14) and (DistinctIntervals >= 25) and (TrackNormalizedDuration >= 0.95) then IsMelody=true
R6	if (AvgAbsInterval <= 2.44) and (TrackNumNotes >= 130) and (AvgPitch >= 55.2) and (TrackOccupationRate >= 0.31) and (TrackPolyphonyRate <= 0.001) then IsMelody=true

C. From crisp to fuzzy rule system

Although informative, this rule system is not easily readable or even understandable at first sight, at least for people as musicians or musicologists. Also, being *melody* such a vague concept, the authors find that a fuzzy description of melody would be more sensible in the imprecise domain of music characterization.

In order to produce such a fuzzy description, a fuzzyfication process is applied to a crisp rule system, such the one presented in Table II.

Two basic steps must be carried out for the fuzzyfication of the crisp rule system. First, the data representation must be fuzzified. That is, numerical input and output values must be converted to fuzzy terms. Second, the rules themselves must be translated into fuzzy rules, substituting linguistic terms for numerical boundaries.

D. Fuzzyfying attributes

As stated above, a MIDI track is described by a set of statistical descriptors (called attributes from herein). The very first step of the attribute fuzzyfication process is to define the domain for every attribute. Most attributes have a finite domain. For practical application of the fuzzification method, infinite domains should be converted to finite domains. Appropriate upper and lower bounds are so defined for these domains.

In order to fuzzify crisp attributes (statistical descriptors), linguistic terms (such as *low*, *average*, or *high*) for every attribute domain are defined. Then the shape of the fuzzy set associated with each linguistic term is selected and, finally, the value of each fuzzy set parameter within the attribute domain is set.

Fuzzyfication of numerical attributes usually involves the participation of a human expert who provides domain knowledge for every attribute. The expert usually takes into consideration the distribution of values for an attribute in a reference data collection, as well as any other information available.

Our approach in this paper is to replace the human expert by a genetic algorithm (GA) which, given the linguistic term definitions for each attribute, automatically learns the fuzzy set parameters. Such combination of a fuzzy system with a genetic algorithm is known as a *genetic fuzzy system* [18].

In order to select the number of linguistic terms per attribute, a number of different crisp rule systems have been induced by different algorithms from the SMALL dataset. The presence of each attribute in those rule systems has been accounted for. Five terms have been assigned to most frequently used attributes. Three terms have been assigned to the rest of attributes. Table III shows these linguistic terms for attributes used in the RIPPER-SMALL crisp rule system.

TABLE III
FUZZY LINGUISTIC TERMS

Attribute	Linguistic terms
TrackNormalizedDuration	*shortest, average, largest*
TrackNumNotes	*low, average, high*
TrackOccupationRate	*void, low, average, high, full*
TrackPolyphonyRate	*none, low, average, high, all*
LowestNormalizedPitch	*low, average, high*
AvgPitch	*veryLow, low, average, high, veryHigh*
StdDeviationPitch	*low, average, high*
DistinctIntervals	*few, average, alot*
NormalizedDistinctIntv.	*lowest, average, highest*
AvgAbsInterval	*unison, second, third, fourth, high*
ShortestDuration	*low, average, high*
ShortestNormalizedDur.	*shortest, average, longest*
TrackSyncopation	*few, average, alot*

Every linguistic term has a fuzzy set or membership function associated to it. This is a probability function from the attribute crisp input domain to the range $[0, 1]$ that, for every possible attribute crisp value, outputs the probability for this value to be named with that specific linguistic term. Figure 1 shows an example.

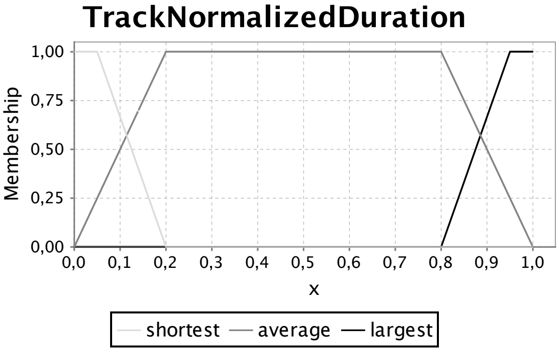


Fig. 1. Fuzzy set example for attribute *TrackNormalizedDuration*

For efficiency reasons, the shape for a fuzzy set in this work is restricted to be either trapezoidal or triangular, being the latter a special case of the former. Each fuzzy set is modeled by four points, corresponding to the extreme points of the *core* (*prototype*) and *support* of a fuzzy set, as depicted in Fig. 2. The support of a fuzzy set defines the range of the input domain where the fuzzy set membership probability is not zero. These fuzzy set parameters would be inferred from data by the GA.

The objective for the genetic fuzzy system presented here is to optimize fuzzy set parameters for every attribute in a fuzzy rule system. This optimization process is guided by a fitness function that, given a reference fuzzy rule system, tests potential solutions against a reference dataset.

1) Fuzzy set representation scheme: An individual's chromosome encodes all attributes of the fuzzy rule system. This means to encode fuzzy sets associated with linguistic terms for every attribute. The fuzzy set support is considered the most important part of a fuzzy set, while its shape is considered a subjective and application-dependent issue [19]. The fuzzy set core is defined as a function of its support. So, the only fuzzy set parameters we need to optimize are the support points of each fuzzy set for every attribute. Figure 3a shows how an attribute domain is partitioned in overlapping fuzzy partitions, each corresponding to a fuzzy set. Let X be such attribute domain, we define

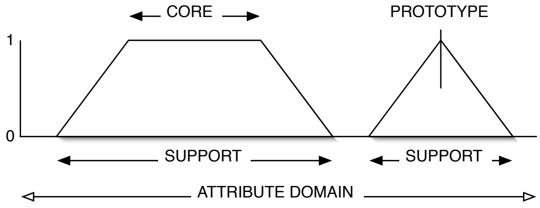


Fig. 2. Fuzzy set parts

a fuzzy partition of X as

$$X^i = \left[x_L^i, x_R^i\right], X^i \subset X, \quad 1 \leqslant i \leqslant m \tag{1}$$

where x_L^i and x_R^i are the *left* and *right support points* of fuzzy set i, respectively. m is the number of fuzzy sets for the attribute. Partitions are defined so that $X = \bigcup X^i$, that is, every input value belong to at least one partition. We also force the overlapping between adjacent partitions i and $i+1$ to be not void:

$$Z^{i,i+1} = X^i \bigcap X^{i+1} = \left[x_L^{i+1}, x_R^i\right] \neq \emptyset \tag{2}$$

Given these definitions, the set of parameters to optimize for a given attribute is

$$\Theta = \{x_L^1, x_L^2, x_R^1, \cdots, x_L^m, x_R^{m-1}, x_R^m\} \tag{3}$$

In order to have an uniform GA representation for every attribute, their domains are normalized to range $[0,1]$, so every parameter is a value in that range. For the sake of simplicity, let express Θ as

$$\Theta = \left\{p_0, p_1, p_2, \cdots, p_{2m-1}\right\} \tag{4}$$

From the partitioning scheme definition, it follows that $p_0 = x_L^1 = 0$, so we can drop this first parameter. In order to make Θ suitable to crossover and mutation operations, a relative parameter representation scheme is used in the GA. Such scheme is defined as follows

$$\theta = \left\{p_1, r_2, r_3, \cdots, r_{2m-1}\right\} \tag{5}$$

where $r_i = p_i - p_{i-1}$. Figure 4 depicts the representation scheme used in the GA. Note that

$$Z^{i,i+1} = r_{2i}, \quad 1 \leqslant i < m$$

.

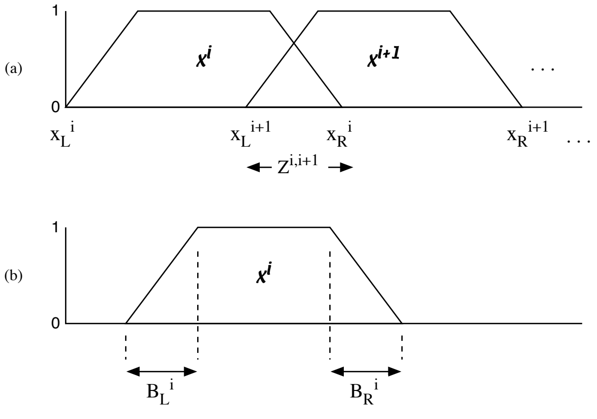


Fig. 3. (a) Fuzzy set partitions overlapping. (b) Boundaries of a fuzzy set.

Once the support points are known, left and right boundaries (figure 3b) are set. They are restricted to lie inside the overlapping section of their corresponding partition. For right boundaries,

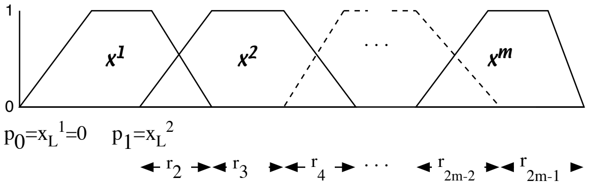


Fig. 4. Representation scheme of fuzzy sets.

$$0 \leqslant B_R^i \leqslant Z^{i,i+1} = r_{2i}, \quad 1 \leqslant i < m$$

and $0 \leqslant B_R^m \leqslant r_{2m-1}$.

For left boundaries

$$0 \leqslant B_L^i \leqslant Z^{i-1,i} = r_{2i-2}, \quad 1 < i \leqslant m$$

and $0 \leqslant B_L^1 \leqslant p_1$. This ensures that the core of a fuzzy set is equal or greater than zero.

E. Fitness function

The fitness function for the GA consists of testing each individual in a fuzzy inference system (FIS) using the fuzzy rule system discussed in section II-F on a reference dataset (see section III). The better the performance of the rule system, given the fuzzy set definitions provided by the individual's chromosome, the better the individual's score. This is possible because rule fuzzification is a process independent from fuzzy set definition. Several metrics can be used to measure the performance of the FIS. In this work two different metrics have been tested: 1) number of hits and 2) F measure (geometric mean of precision and recall of class *IsMelody=true*).

F. Crisp rule system fuzzyfication

The goal of the rule system presented above is to identify MIDI tracks as melody or non-melody tracks. The objective of this work is to convert this crisp rule system, which perform fairly well for the task at hand, in a human-friendly description of melody tracks.

The final step in this method is to fuzzify the rule system. Antecedents of the form $(x \circledcirc v)$ where $\circledcirc$ is an inequality operator, are translated into one or more antecedents of the form $(x\ IS\ T)$, where T is a linguistic term defined for attribute x. The value v partitions the attribute domain in two subsets, and the direction of the inequality guides the selection of the fuzzy terms to be included in fuzzy antecedents.

In the present work, the crisp RIPPER-SMALL rule system (section II-B) has been fuzziyfied in order to present a proof of concept of the methodology applied. A disjunctive fuzzy rule set is then obtained. Table IX shows fuzzy rules corresponding to those shown in section II-B.

III. Experiments

A. Datasets

Table IV shows information about all the datasets used to test the fuzzy rule system. They consist of MIDI files, where melody tracks were tagged with a special string in their track name. These tracks have been manually or automatically tagged, depending on the dataset. The automatic tagging process is based on a dictionary of frequent melody track names. The manual tagging was carried out by experts on the different music genres present in the datasets.

The SMALL reference dataset has been used to obtain the crisp rule system from which the fuzzy rule system has been derived. It is also the dataset used in the GA fitness function to test the performance of potential solutions. The rest of datasets are used for testing the system: RWC-G [20], RWC-P [21], LARGE and AJP are all multi-genre datasets of academic, popular, rock and jazz music, among more than ten genres.

TABLE IV

Datasets.

Dataset	Tracks	Songs	Melody tracks
SMALL	2775	600	554
LARGE	15168	2513	2337
RWC-P	801	75	74
RWC-G	311	48	44
AJP	3732	762	760

B. FIS Optimization Experiment setup

Our genetic fuzzy system has six free parameters that let configure different experiment setups. Table V shows these parameters and the values chosen to build a set of experiments. Parameter values have been restricted to at most three different values. This allows the use of an orthogonal array [22] to explore the free parameter space. Briefly, an orthogonal array of level L, strength n and M runs ensures that, given any n parameters with L values each, all their respective values will appear in combination in an equal number of experiments. This avoids testing all possible combinations, while remaining confident that every combination of n parameters appears at least once in some experiment. In this work, an orthogonal array of strength 2 and 18 runs has been used to setup experiments.

TABLE V

FIS optimization setup parameters

Experiment parameter	Values
GA population size	100, 500, 1000
GA no. of generations	100, 500, 1000
GA mutation ratio	none, 0.05, 0.1
GA selection strategy[2]	Best one, Best 10%, Best 20%
GA fitness metric	Hit count, F-measure
Defuzzyfication threshold[3]	0.5, 0.6, 0.7

IV. Fuzzy Inference System Optimization Results

Table VI shows the performance of evolved FIS versus the RIPPER-SMALL crisp rule system performance. Average results from the eighteen experiments performed are shown. Figures in parenthesis are standard deviations. Precision, recall and F-measure are computed for the class 'IsMelody'. Also, the performance of the best evolved FIS are presented. Note that the best evolved FIS performance is very close to that from the crisp rule system. The definition of fuzzy sets for the best evolved FIS, as well as other information and examples on this work can be found on the web at the following address: *http://grfia.dlsi.ua.es/cm/worklines/smc08*.

TABLE VI

Best and average performance of evolved FIS v. crisp RIPPER-SMALL rule system performance.

Rule sys.	Precision	Recall	F	Error rate
crisp	0.89	0.87	0.88	0.05
Best FIS	0.81	0.83	0.82	0.06
Avg. FIS	0.80 (.03)	0.77 (.09)	0.78 (.05)	0.08 (.01)

V. Results on test datasets.

Table VII presents results from applying both the crisp rule system and the best evolved FIS to test datasets. In these test experiments, a track is classified as a melody track if it fires at least one rule with probability greater than 0.5. Otherwise, the track is classified as non-melody.

TABLE VII

Melody track classification results.

Dataset	Precision	Recall	F	Error rate
LARGE (*crisp*)	0.79	0.80	0.80	0.06
LARGE (*fuzzy*)	0.70	0.74	0.72	0.09
RWC-P (*crisp*)	0.95	0.80	0.87	0.02
RWC-P (*fuzzy*)	0.51	0.64	0.57	0.09
RWC-G (*crisp*)	0.54	0.77	0.64	0.13
RWC-G (*fuzzy*)	0.43	0.43	0.43	0.16
AJP (*crisp*)	0.88	0.89	0.88	0.05
AJP (*fuzzy*)	0.88	0.83	0.86	0.06

As the results show, the fuzzyfied rule system precision is consistenty lower than the precision of the original crisp rule system. The bigest differences in precision between the fuzzy and crisp rule systems is observed in the smallest data sets, i.e. RWC-P AND RWC-G, with a limited set of examples (e.g. RWC-G contains only 44 melody examples). However, in the LARGE and AJP data sets the difference in precisions of the two rule systems is less considerable. The recall is consistently better for the fuzzy classifier. It follows that most errors are false positives, that is, some non-melody tracks are classified as melody tracks. Also note that the goal of the fuzzyfication process is not to improve classification accuracy, but to obtain a human-readable comprehensible characterization of melodies within MIDI tracks.

VI. Comparison of crisp and fuzzy systems on some examples

This section discuss several example characterization of melody and non-melody tracks. The example excerpts are shown in Table VIII in the appendix. The words 'Crisp' and 'Fuzzy' under the music systems indicate which rules from the crisp and fuzzy systems were fired, respectively. The fuzzy rule system used with these examples was the best evolved FIS using the rules in Table IX.

The first three tracks are melody tracks that were correctly identified by the fuzzy rule system. Crisp rules failed at characterizing the first one. This first track almost fulfills rule $R2$, except that it has not the largest pitch interval variety (its *NormalizedDistinctIntervals* value is .85), as the last condition of the rule imposes. The next three tracks in Table VIII are non-melody tracks correctly identified by both rule systems (neither track fire any rule). The last two examples are tracks were both rule systems disagree. The melody track from *Satin Doll* is unusual in the senese that is supposed to be played by a vibraphone (a polyphonic instrument), has one chorus of improvisation and the melody reprise (which is the part shown in the example) is played in a polyphonic *closed chord* style. The last example is a piano accompaniment part, played in *arpeggiato* style, which the fuzzy rules incorrectly identified as a melody track. This track almost fired crisp rule $R6$, except for the last condition of the rule, because its *TrackPolyphonyRate* value is .097. This is a clear example of why a fuzzy version of a crisp rule fires while the crisp rule don't. The value is accepted by the fuzzy rule as linguistic term *none* for the *TrackPolyphonyRate* attribute. This is because it lies into the support of the fuzzy set corresponding to that term. See figure 5 for some fuzzy set examples from the best evolved FIS.

VII. Conclusions and further work

We presented an approach to automatic human-readable melody characterization using fuzzy rules. We considered MIDI files, and extracted a set of statistical descriptors from MIDI files datasets. We then applied a rule induction algorithm to obtain a set of (crisp) classification rules for melody track identification. Finally, we automatically transformed the crisp rules into fuzzy rules by applying a genetic algorithm to generate the membership functions for the rule attributes. The classification accuracy of the resulting fuzzy rule system is lower than the original crisp rule system, but comprehensibility of the rues is improved. We plan to improve the performance of the fuzzy rule system by modifying (i.e. rising) the probability threshold for firing a fuzzy rule. Also, enforcing more than one fuzzy rule to be fired could help improve the results. We plan to explore alternative approaches for the rule fuzzyfication, e.g. by using information theory measures.

Acknowledgments

The authors want to thank Pablo Cingolani and the rest of contributors to the *jFuzzyLogic* package (http://jfuzzylogic.sourceforge.net) we used to implement our fuzzy rule systems. Also we want to thank Klaus Meffert et al. as major contributors to the *Java Genetic Algorithms Package, JGAP* (http://jgap.sourceforge.net), the one we used to implement our GA experiments. Last, but not least, thanks to the people at University of Waikato behind the *weka* project (http://www.cs.waikato.ac.nz/ml/weka/), used to build our crisp rule models.

This work is supported by the projects: GV06/166 and CICyT TIN2006–14932–C02, partially supported by EU ERDF.

References

[1] A. Ghias, J. Logan, D. Chamberlin, and B. C. Smith, "Query by humming: Musical information retrieval in an audio database," in *Proc. of 3rd ACM Int. Conf. Multimedia*, 1995, pp. 231–236.

[2] J. Eggink and G. J. Brown, "Extracting melody lines from complex audio," in *ISMIR*, 2004.

[3] I.Karydis, A.Nanopoulos, A.Papadopoulos, E. Cambouropoulos, and Y. Manolopoulos, "Horizontal and vertical integration/segregation in auditory streaming: a voice separation algorithm for symbolic musical data," in *Proceedings 4th Sound and Music Computing Conference (SMC'2007)*, Lefkada, 2007.

[4] M. Tang, C. L. Yip, and B. Kao, "Selection of melody lines for music databases." in *Proceedings of Annual Int. Computer Software and Applications Conf. COMPSAC*, 2000, pp. 243–248.

[5] S. T. Madsen and G. Widmer, "Towards a computational model of melody identification in polyphonic music." in *20th International Joint Conference on Artificial Intelligence (IJCAI 2007)*, 2007, pp. 459–464.

[6] E. Toch, *La melodía (translation of 'Melodielehre', 1923)*. Span-Press Universitaria, 1997.

[7] S. Sadie and G. Grove, *The New Grove Dictionary of Music and Musicians*. Macmillan, 1084.

[8] E. Selfridge-Field, *Conceptual and representational issues in melodic comparison*, ser. Computing in Musicology. Cambridge, Massachusetts: MIT Press, 1998, vol. 11, pp. 3–64.

[9] M.Baroni, *Proposal for a Grammar of Melody: The Bach Chorales*. Les Presses de l'Université de Montréal, 1978.

[10] D. Cope, *Experiments in Musical Intelligence*. New York, NY, USA: Cambridge University Press, 1996, vol. 2, no. 1.

[11] E.Narmour, *The Analysis and Cognition of Basic Melodic Structures*. University Of Chicago Press, 1990.

[12] Y. E. Kim, W. Chai, R. Garcia, and B. Vercoe, "Analysis of a contour-based representation for melody," in *ISMIR*, 2000.

[13] A. E. Gomez, A. Klapuri and B.Meudic, "Melody description and extraction in the context of music content processing," *Journal of New Music Research (JNMR)*, vol. 32-1, 2003.

[14] D. Temperley, *The Cognition of Basic Musical Structures*. The MIT Press, 2004.

[15] P. J. Ponce de León, D. Rizo, and J. M. Iñesta, "Towards a human-friendly melody characterization by automatically induced rules," in *Proceedings of the 8th International Conference on Music Information Retrieval*, S. Dixon, D. Brainbridge, and R. Typke, Eds. Vienna: Austrian Computer Society, September 2007, pp. 437–440.

[16] D. Rizo, P. J. Ponce de León, C. Pérez-Sancho, A. Pertusa, and J. M. Iñesta, "A pattern recognition approach for melody track selection in midi files," in *Proc. of the 7th Int. Symp. on Music Information Retrieval ISMIR 2006*, T. A. Dannenberg R., Lemström K., Ed., Victoria, Canada, 2006, pp. 61–66, iSBN: 1-55058-349-2.

[17] W. W. Cohen, "Fast effective rule induction," *Machine Learning: Proceedings of the Twelfth International Conference*, 1995.

[18] O. Cordón and F. Herrera, "A general study on genetic fuzzy systems," in *Genetic Algorithms in Engineering and Computer Science*, J. Smith, Ed. John Wiley & Sons, 1995, ch. 3, pp. 33–57.

[19] M. Makrehchi, O. A. Basir, and M. Kamel, “Generation of fuzzy membership function using information theory measures and genetic algorithm,” in *Fuzzy Sets and Systems - IFSA 2003*, ser. Lecture Notes in Computer Science, T. Bilgiç, B. D. Baets, and O. Kaynak, Eds., vol. 2715. Springer, 2003, pp. 603–610.

[20] M. Goto, H. Hashiguchi, T. Nishimura, and R. Oka, “RWC music database: Music genre database and musical instrument sound database.” in *ISMIR*, 2003.

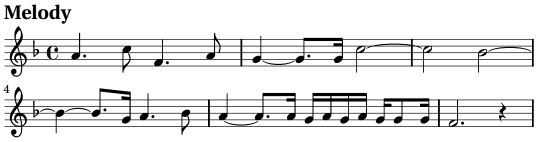
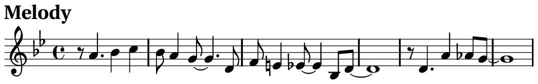
[21] ——, “RWC music database: Popular, classical and jazz music databases.” in *ISMIR*, 2002.

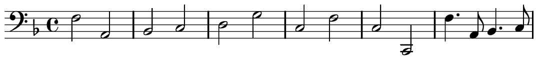
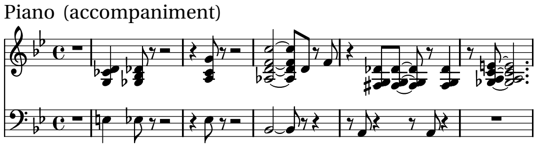
[22] A. Hedayat, N. J. A. Sloane, and J. Stufken, *Orthogonal Arrays: Theory and Applications*, 1st ed. Springer, 1999.

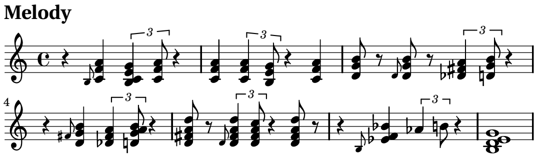
Appendix

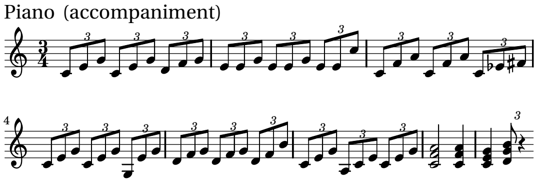
TABLE VIII

Track classification examples.

True positive examples
Air In F, Watermusic, Handel (*Baroque*) **Melody**  Crisp: – Fuzzy: FR6
There Is No Greater Love, I. Jones (*pre-Bop Jazz*) **Melody**  Crisp: R2, R5 Fuzzy: FR4, FR6

True negative examples
Air In F, Watermusic Bass  Crisp: – Fuzzy: –
There Is No Greater Love Piano (accompaniment)  Crisp: – Fuzzy: –

False negative example
Satin Doll, D. Ellington (*pre-Bop Jazz*) **Melody**  Crisp: R2 Fuzzy: –

False positive example
Sonata no. 3 K545, 2nd Mov., W.A. Mozart (*Classicism*) Piano (accompaniment)  Crisp: – Fuzzy: FR6

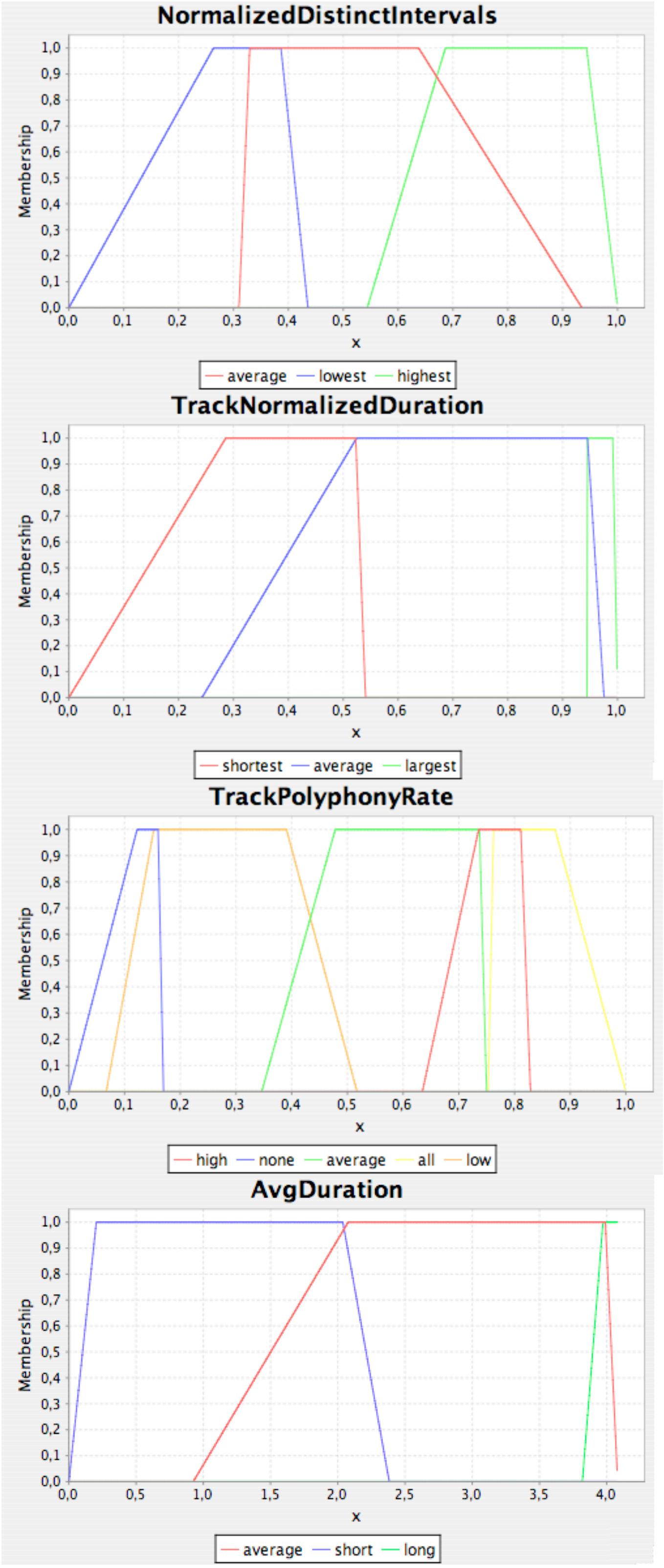


Fig. 5. Fuzzy set examples from the best evolved fuzzy rule system.

TABLE IX

FUZZY RULES EQUIVALENT TO THOSE SHOWN IN TABLE II

Name	Rule	Name	Rule
FR1	IF (AvgPitch IS high OR AvgPitch IS veryHigh) AND (TrackOccupationRate IS NOT void) AND (TrackOccupationRate IS NOT low) AND (AvgAbsInterval IS NOT fourth) AND (AvgAbsInterval IS NOT high) AND (TrackNumNotes IS high) THEN IsMelody IS true	FR2	IF (AvgPitch IS high OR AvgPitch IS veryHigh) AND (TrackOccupationRate IS NOT void) AND (TrackOccupationRate IS NOT low) AND (TrackPolyphonyRate IS NOT average) AND (TrackPolyphonyRate IS NOT high) AND (TrackPolyphonyRate IS NOT all) AND (NormalizedDistinctIntervals IS highest) THEN IsMelody IS true
FR3	IF (AvgPitch IS high OR AvgPitch IS veryHigh) AND (TrackNumNotes IS high) AND (LowestNormalizedDuration IS shortest) AND (ShortestDuration IS NOT low) AND (NormalizedDistinctIntervals IS highest) THEN IsMelody IS true	FR4	IF (AvgPitch IS high OR AvgPitch IS veryHigh) AND (TrackOccupationRate IS NOT void) AND (TrackOccupationRate IS NOT low) AND (AvgAbsInterval IS NOT third) AND (AvgAbsInterval IS NOT fourth) AND (AvgAbsInterval IS NOT high) AND (TrackSyncopation IS NOT few) AND (StdDeviationPitch IS NOT high) THEN IsMelody IS true
FR5	IF (AvgAbsInterval IS NOT fourth) AND (AvgAbsInterval IS NOT high) AND (TrackSyncopation IS alot) AND (LowestNormalizedPitch IS NOT low) AND (DistinctIntervals IS alot) AND (TrackNormalizedDuration IS largest) THEN IsMelody IS true	FR6	IF (AvgPitch IS NOT veryLow) AND (AvgPitch IS NOT low) AND (TrackOccupationRate IS NOT void) AND (TrackOccupationRate IS NOT low) AND (AvgAbsInterval IS NOT third) AND (AvgAbsInterval IS NOT fourth) AND (AvgAbsInterval IS NOT high) AND (TrackPolyphonyRate IS none) AND (TrackNumNotes IS NOT low) THEN IsMelody IS true

Expressive Performance in the Human Tenor Voice

Maria Cristina Marinescu*, Rafael Ramirez†
*IBM T.J. Watson Research Center/USA
†Universitat Pompeu Fabra/Barcelona, Spain

***Abstract*—This paper presents preliminary results on expressive performance in the human tenor voice. This work investigates how professional opera singers manipulate sound properties such as timing, amplitude, and pitch in order to produce expressive performances. We also consider the contribution of features of prosody in the artistic delivery of an operatic aria. Our approach is based on applying machine learning to extract patterns of expressive singing from performances by Josep Carreras. This is a step towards recognizing performers by their singing style, capturing some of the aspects which make two performances of the same piece sound different, and understanding whether there exists a correlation between the occurrences correctly covered by a pattern and specific emotional attributes.**

I. Introduction

One of the most interesting and elusive questions in music is what makes two expressive interpretations of the same musical piece sound like two different songs even when performed by the same singer. Given a set of expressive performances of the same piece which have different interpretation styles — and possibly different emotional attributes[1] — are the patterns learned from each performance similar or very different? How distinguishable is a singer based on the patterns extracted from his interpretations? What do the patterns that are similar for multiple singers capture? Which patterns are a matter of timber, which are based in specific expressive techniques that a singer employs, and which are a combination of the two — by choice or because the pattern is more readily realizable given the characteristics of a specific voice? Is there a correlation between the occurrences correctly covered by a pattern and specific emotional attributes associated with those music pieces? How do singers resolve possible conflicts between the music and the prosody of the lyrics?

This work investigates how professional opera singers manipulate sound properties such as timing, amplitude, and pitch in order to produce expressive performances of music fragments. In the initial phase we are interested in note-level manipulations; we therefore define a set of note-level descriptors of interest and we focus on the differences between their measured values in the actual performance and the written score, given the context of the surrounding notes. Previous approaches exists that are looking at expressive instrumental performances. There are a couple of important differences between instrumental music and voice in operatic music. First, one guitar may subtly differ from another one, but the timbre of the instrument is relatively fixed. Human voice displays a great variety in timbre; it is partly because of timbre that a voice is well-suited to a type of song but not to another, and it may be because of our preference in timbre that we prefer a singer over another. Distinguishing which features of expressiveness are the singer's interpretation choice and which ones are typical of his timbre is an issue that doesn't come up in instrumental music.

Secondly, instrumental music does not have lyrics. Lyrics convey a more specific meaning to a song than it would otherwise have. Therefore they can both add to and detract from the expressivity of a performance. Several aspects are at work here: how appropriate in meaning is the performance given the lyrics, and how to reconciliate possibly contradicting prosodic, metric, and score cues. For instance, adopting the wrong intonation or grouping the lyrics into the wrong prosodic units can ruin an otherwise good interpretation. In this work we are looking at a couple of preliminary descriptors for the lyrics which are syllable-specific: stress and syllable type.

Our approach is based on applying various machine learning (ML) techniques to extract patterns of expressive singing from different performances of the same, or different arias, sung by several world-class tenors. As a first step, we start with a test suite consisting of twelve interpretations of six different aria fragments performed by Josep Carreras. Using sound analysis techniques based on spectral models we extract high-level descriptors representing properties of each note, as well as of its context. A note is characterized by its pitch and duration. The context information for a given note consists of the relative pitch and duration of the neighbouring notes, as well as the Narmour structures to which the note belongs. In this work, our goal is to learn under which conditions a performer shortens or lengthens a note relative to what the score indicates, and when he sings a note louder or softer than what would be expected given the average energy level of the music fragment. Some of the most interesting rules that the ML algorithm learns are presented in the result section.

The rest of the paper is organized as follows. Section II describes related work in expressive performance. Section III describes our test suite, introduces the note-level descriptors, and explains how we extract the data that

[1] Emotional attributes are similar to what other researchers refer to as *moods* or *affective labels* and can simultaneously take one value for each aspect that they reflect.

is used as the input to the ML algorithms. Section IV presents the learning algorithms; Section V discusses some of the most interesting results. We conclude in Section VI.

II. Related Work

Understanding and formalizing expressive music performance is an extremely challenging problem which in the past has been studied from different perspectives, e.g. [16], [6], [3]. The main approaches to empirically studying expressive performance have been based on statistical analysis (e.g. [15]), mathematical modeling (e.g. [19]), and analysis-by-synthesis (e.g. [5]). In all these approaches, it is a person who is responsible for devising a theory or mathematical model which captures different aspects of musical expressive performance. The theory or model is later tested on real performance data in order to determine its accuracy. This paper describes a machine learning approach to investigate how opera singers express and communicate their view of the musical and emotional content of musical pieces.

Previous research addressing expressive music performance using machine learning techniques has included a number of approaches. Widmer [20] reported on the task of discovering general rules of expressive classical piano performance from real performance data via inductive machine learning. The performance data used for the study are MIDI recordings of 13 piano sonatas by W.A. Mozart performed by a skilled pianist. In addition to these data, the music score was also coded. The resulting substantial data consists of information about the nominal note onsets, duration, metrical information and annotations. When trained on the data an inductive rule learning algorithm discovered a small set of quite simple classification rules that predict a large number of the note-level choices of the pianist.

Tobudic et al. [18] describe a relational instance-based approach to the problem of learning to apply expressive tempo and dynamics variations to a piece of classical music, at different levels of the phrase hierarchy. The different phrases of a piece and the relations among them are represented in first-order logic. The description of the musical scores through predicates (e.g. *contains(ph1,ph2)*) provides the background knowledge. The training examples are encoded by another predicate whose arguments encode information about the way the phrase was played by the musician. Their learning algorithm recognizes similar phrases from the training set and applies their expressive patterns to a new piece.

Ramirez et al. [13], [14] explore and compare different machine learning techniques for inducing both, an *interpretable* expressive performance model (characterized by a set of rules) and a *generative* expressive performance model. Based on this, they describe a performance system capable of generating expressive monophonic Jazz performances and providing 'explanations' of the expressive transformations it performs. The work described in this chapter has similar objectives but by using a genetic algorithm it incorporates some desirable properties: (1) the induced model may be explored and analyzed while it is 'evolving', (2) it is possible to guide the evolution of the model in a natural way, and (3) by repeatedly executing the algorithm different models are obtained. In the context of expressive music performance modeling, these properties are very relevant.

Lopez de Mantaras et al. [8] report on SaxEx, a performance system capable of generating expressive solo performances in jazz. Their system is based on case-based reasoning, a type of analogical reasoning where problems are solved by reusing the solutions of similar, previously solved problems. In order to generate expressive solo performances, the case-based reasoning system retrieves, from a memory containing expressive interpretations, those notes that are *similar* to the input inexpressive notes. The case memory contains information about metrical strength, note duration, and so on, and uses this information to retrieve the appropriate notes. However, their system does not allow one to examine or understand the way it makes predictions.

Other inductive machine learning approaches to rule learning in music and musical analysis include [4], [1], [9] and [7]. In [4], Dovey analyzes piano performances of Rachmaninoff pieces using inductive logic programming and extracts rules underlying them. In [1], Van Baelen extended Dovey's work and attempted to discover regularities that could be used to generate MIDI information derived from the musical analysis of the piece. In [9], Morales reports research on learning counterpoint rules. The goal of the reported system is to obtain standard counterpoint rules from examples of counterpoint music pieces and basic musical knowledge from traditional music. In [7], Igarashi et al. describe the analysis of respiration during musical performance by inductive logic programming. Using a respiration sensor, respiration during cello performance was measured and rules were extracted from the data together with musical/performance knowledge such as harmonic progression and bowing direction.

III. Expressive singing in the tenor voice

Our choice of studying the human singing voice in the operatic context is not arbitrary; in fact, we believe that operatic music is an ideal environment to start getting some answers to our questions. First, there is a constrained environment in which the music is performed and which is given by the written score and the meaning of the lyrics. Keeping such variables fixed makes the results and comparisons between different singers more meaningful. It also makes it easier for a listener to characterize different performances from the point of view of their emotional attributes. Secondly, good operatic singers tend to have both better voice and better technique than singers in most other genre, and can employ them more efficiently for expressive interpretations. In this context, we choose to focus on the most sought-after role in operas, the human tenor voice, arguably the role for which the most famous arias have even been written.

A. Training data

We have chosen six fragments of arias from Rigoletto, Un Ballo in Maschera, and La Traviata. For four of the fragments we have selected two different interpretations; one of the remaining two fragments has three different interpretations, while the remaining one has a single interpretation. In total the twelve fragments consist of 415 notes in which the tenor and the orchestra do not overlap. The choice of interpretations is not random; we have tried to incorporate very different, yet expressive, performances of the same piece. One of the questions we are interested in answering is whether the expressivity patterns we learn from interpretations of the same aria by the same singer are similar despite the different feel of each performance we choose.

One of the reasons we chose to focus on Josep Carreras as a test case is our subjective observation that his interpretations are highly expressive, yet at the same time they can exhibit a wide variation in emotional attributes even over different performances of the same aria. Another reason why he is the ideal candidate for us is that both the timbre of his voice and his delivery have changed considerably over time. In general, we make the assumption that timber does not vary significantly over short periods of time, but it may change dramatically over long periods. By studying recordings that are close in time we can compare expressivity patterns while controlling over the timbre. Studying recordings that are chronologically far but exhibit the same emotional attributes can on the other hand help understanding which of the patterns we learn are greatly affected by changes in timbre and which are not. We therefore keep track of the recording date of the interpretations that we are processing.

A secondary reason to record this information has to do with what we call *appropriateness* of an interpretation — the capacity of a singer to inhabit a musical piece. Defining this measure is an interesting topic in itself, and touches on many aspects including the question of meaning in music. Our assumption is that recordings closer in time of arias sung in a language familiar to the tenor will minimize appropriateness variations. Future experiments aim to selectively control over the effect of such factors.

B. Musical analysis

We use sound analysis techniques based on spectral models [17] for extracting high-level symbolic features from the recordings. We characterize each performed note by a set of features representing both properties of the note itself and aspects of the musical context in which the note appears. Information about the note includes note pitch and note duration, while information about its melodic context includes the relative pitch and duration of the neighboring notes (i.e. previous and following notes) as well as the Narmour structures to which the note belongs.

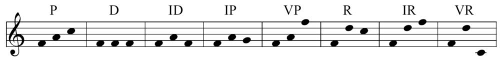


Fig. 1. Prototypical Narmour structures

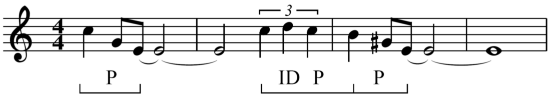


Fig. 2. Narmour analysis of a musical fragment

In order to provide an abstract structure to our performance data, we decided to use Narmours theory [10] to analyze the performances. The Implication/Realization model proposed by Narmour is a theory of perception and cognition of melodies. The theory states that a melodic musical line continuously causes listeners to generate expectations of how the melody should continue. According to Narmour, any two consecutively perceived notes constitute a melodic interval, and if this interval is not conceived as complete, it is an implicative interval, i.e. an interval that implies a subsequent interval with certain characteristics. That is to say, some notes are more likely than others to follow the implicative interval. Based on this, melodic patterns or groups can be identified that either satisfy or violate the implication as predicted by the intervals. Figure 1 shows prototypical Narmour structures. We parse each melody in the training data in order to automatically generate an implication/realization analysis of the pieces. Figure 2 shows the analysis for a fragment of *All of me*.

We additionally annotate the lyrics with syllable-specific information. In our fragments it is overwhelmingly the case that a syllable corresponds to a note in the score. The exceptions are few; in one instance two syllables of a word correspond to a single note. The rest of the ten cases are instances in which the last syllable of a word ends in a vowel and the first syllable of the following one starts with a vowel and they together correspond to a single note in the score. For the beginning we simply specify which syllables are stressed or unstressed, and whether they are open or closed. The librettos for all the fragments in the test suite are written in Italian. If any of the syllables which correspond to a note is stressed then the note will be stressed. In Italian a syllable is open if it ends in a vowel and closed otherwise.

Lastly, we want to see how prosody interacts with the score and the meter of the lyrics. We consider that prosody can give important clues about the emotional content that the singer wants to communicate as it reflects aspects that are not inherent in the lyrics: intonation, rhythm, and 'prosodic' stress. For instance, many have observed that stress may be a matter of the prosodic unit rather than the actual stress of the words. A prosodic unit is a unit of meaning which can be as short as a word and as long as a statement; it is a chunk of speech that may in fact reflect how the brain processes speech. Acoustically, a prosodic unit is characterized by a few phonetic cues: (1) a typical pitch contour which gradually declines towards the end of the unit and resets itself at the beginning of the next unit, (2) perceptual discontinuities between units, (3) long final unit words. We are interested in where the actual stress falls in a performance, which syllables are over-articulated, what the pitch contour can tell us about the emotional state that the singer transmits, and how are potential conflicts solved between the stress in a prosodic unit and the meter of the lyrics. To make such observations we need to (1) establish the meter of the lyrics and (2) split the lyrics into prosodic units.

C. Learning task

For each expressive transformation, we approach the problem both as a regression and a classification problem. As a regression problem we learn a model for predicting the lengthening ratio of the performed note wrt the score note. This is, a predicted ratio greater than 1 corresponds to a performed note longer than as specified in the score, while a predicted ration smaller than 1 coresponds to a shortened performed note (e.g. a 1.15 prediction corresponds to a 15% performed note lengthening wrt the score). As a classification problem, the performance classes of interest are *lengthen, shorten* and *same* for duration transformation, and *soft, loud* and *same* for energy variation. A note is considered to belong to class *lengthen*, if its performed duration is 20% longer (or more) that its nominal duration, e.g. its duration according to the score. Class *shorten* is defined analogously. A note is considered to be in class *loud* if it is played louder than its predecessor and louder than the average level of the piece. Class *soft* is defined analogously. We decided to set these boundaries after experimenting with different ratios. The main idea was to guarantee that a note classified, for instance, as lengthen was purposely lengthened by the performer and not the result of a performance inexactitude.

IV. Learning Algorithm

We used Tilde's top-down decision tree induction algorithm [2]. Tilde can be considered as a first order logic extension of the C4.5 decision tree algorithm: instead of testing attribute values at the nodes of the tree, Tilde tests logical predicates. This provides the advantages of both propositional decision trees (i.e. efficiency and pruning techniques) and the use of first order logic (i.e. increased expressiveness). The increased expressiveness of first order logic not only provides a more elegant and efficient specification of the musical context of a note, but it provides a more accurate predictive model [12].

We apply the learning algorithm with target predicates: `duration/3` and `energy/3`. (where `/n` at the end of the predicate name refers to the predicate arity, i.e. the number of arguments the predicate takes). Each target predicate corresponds to a particular type of transformation: `duration/3` refers to duration transformation and `energy/3` to energy transformation.

For each target predicate we use as example set the complete training data specialized for the particular type of transformation, e.g. for `duration/3` we used the complete data set information on duration transformation (i.e. the performed duration transformation for each note in the data set). The arguments are the musical piece, the note in the piece and performed transformation.

We use (background) predicates to specify both note musical context and background information. The predicates we consider include `context/8`, `narmour/2`, `succ/2` and `member/3`. Predicate `context/8` specifies the local context of a note. i.e. its arguments are *(Note,Pitch, Dur, MetrStr, PrevPitch, PrevDur, NextPitch, NextDur)*. Predicate `narmour/2` specifies the Narmour groups to which the note belongs. Its arguments are the note identifier and a list of Narmour groups. Predicate `succ(X,Y)` means `Y` is the successor of `X`, and Predicate `member(X,L)` means `X` is a member of list `L`. Note that `succ(X,Y)` also means that `X` is the predecessor of `Y`. The `succ(X,Y)` predicate allows the specification of arbitrary-size note-context by chaining a number of successive notes:

$$succ(X_1, X_2),\ succ(X_2, X_3), \ldots,\ succ(X_{n-1}, X_n)$$

where X_i $(1 \leq i \leq n)$ is the note of interest.

V. Results

The induced classification rules are of different types. Both, rules referring to the local context of a note, i.e. rules classifying a note solely in terms of the timing, pitch and metrical strength of the note and its neighbors, as well as compound rules that refer to both the local context and the Narmour structure were discovered. We discovered a few interesting duration rules:

IF Metrical_Strength = veryweak AND
Note_Duration ∈ (-inf, 0.425] AND
Next_Interval ∈ (-1.5, 0.6] AND
Syllable_Stress = stressed
THEN Stretch_Factor = 2.515625

The note duration is measured as the fraction of a beat, where a beat is a quarter note. The interval is measured in number of semitones. The metrical strength is *verystrong* for the first beat, *strong* for the third beat, *medium* for the second and fourth beats, *weak* for the offbeat, and veryweak for any other position of the note. The rule above says that the notes that are in a very weak metrical position, are shorter or equal then 0.425 of a beat (roughly an eight of a note or less), are followed by a note that is lower by at most 1.5 semitones or higher by at most 0.6 semitones, and correspond to a syllable which is stressed, are performed as a 2.5 times longer note than the duration of the note in the score. What is interesting is that a rule with precisely the same *Metrical_Strength*, *Note_Duration*, and *Next_Interval* is performed only 1.3 longer if the corresponding syllable is not stressed.

The next interesting rule has the following form:

IF Metrical_Strength = medium AND

Note_Duration $\in$ *(0.425, 0.6] AND*
Next_Interval $\in$ *(2.7, 4.8] AND*
Syllable_Stress = unstressed AND
narmour(VR, gr_2)
    THEN Stretch_Factor = 2.5

narmour(VR, gr_2) says that the note is in the last (third) position of the registral reversal Narmour structure (VR). Informally this rule says that a note that signals a change of register direction between two intervals of moderate to large size is performed 2.5 longer than the duration of the note in the score if it corresponds to a syllable that is not stressed and it is in the second or fourth beat position. The algorithm also learns two interesting rules about note duration shortening:

IF Metrical_Strength = weak AND
Note_Duration $\in$ *(0.425, 0.6] AND*
Next_Interval $\in$ *(-1.5, 0.6] AND*
Syllable_Stress = stressed AND
narmour(R, gr_2)
    THEN Stretch_Factor = 0.328125

IF Metrical_Strength = weak AND
Note_Duration $\in$ *(0.425, 0.6] AND*
Next_Interval $\in$ *(-1.5, 0.6] AND*
Syllable_Stress = unstressed AND
narmour(P, gr_2)
    THEN Stretch_Factor = 0.40625

These rule indicate that a note corresponding to a stressed syllable immediately following a higher note, and which will be followed by a note close in frequency will be reduced in length to 0.3 of its duration in the score. This technique would accentuate the final note of the largest local ascending interval. Similarly, a small ascending interval that comes after another small interval in the same direction and which corresponds to an unstressed syllable will be shortened to 0.4 of its duration in the score. According to the Narmour principles, a small interval will be followed by another small interval in the same direction; therefore if the note corresponds to a syllable which is not stressed then its importance will be diminished by shortening its duration. On the other hand, if the unstressed note is at the end of a short descending interval followed by a larger descending interval then the note's duration will be lengthened to 1.9 of its duration in the score, in preparation for the downward 'plunge':

IF Metrical_Strength = weak AND
Note_Duration $\in$ *(0.425, 0.6] AND*
Next_Interval $\in$ *(-3.6, -1.5] AND*
Syllable_Stress = unstressed AND
narmour(IP, gr_2)
    THEN Stretch_Factor = 1.90625

An example of energy classification rule is:

IF succ(C, D) AND
narmour(A, D, [nargroup(d, 1)| E]) AND
narmour(A, C, [nargroup(d, 1)| E]) .
THEN energy(A, C, loud) :-

This is, "perform a note loudly if it belongs to an D Narmour group in first position and if its successor belongs to a D Narmour group in first position".

while examples of energy regresion rules are:

IF Note_Duration $\in$ *(0.425, 0.6] AND*
Prev_Interval $\in$ *(-4.8, -2.7] AND*
narmour(IP, gr_1)
    THEN Energy = 109.3799415

That is, "perform a note loudly if it belongs to an IP Narmour group in the second position and if its predecessor interval is a large ascending interval". A similar interpretation has the following rule for a R Narmour group:

IF Note_Duration $\in$ *(0.425, 0.6] AND*
Prev_Interval $\in$ *(-inf, -6.9] AND*
narmour(R, gr_1)
    THEN Energy = 103.715628

Intuitively these two rules say that there is usually a low note that prepares a high, loud note.

A. Prosody vs. Meter

Let us consider one of the three interpretations of the aria **Forse la soglia attinse** from **Un Ballo in Maschera** by Giuseppe Verdi, specifically the recording from 1975 at La Scala. Let us analyze the fragment consisting of *Ah l'ho segnato SILENCE Ah l'ho segnato SILENCE il sacrifizio mio*. There are three prosodic units (PU) here, separated by the silences. The rhythm is iambic. The stress will therefore fall on *l'ho*, *gna*, *sa*, *fi*, and *mi*; these positions are said to be strong and the rest are weak. In the actual interpretation the second "Ah" is stressed, and according to the iambic meter it raises a conflict between the stress of the meter and the prosody. Accentuating a syllable which is in a weak position creates forward motion towards the next stressed syllable in a strong position, namely *gna* (in what is called a *stress valley* [21]). The strong stress on *gna* gives a sense of positive closure. On the other hand the frequency at which the second prosodic unit ends is high (above 300Hz). This is not a typical terminal shape for a prosodic unit as the high pitch suggests something more to come, an arousing rather than settling interest. This is the qualification of the action in PU2 and arrives in form of PU3 — *il sacrifizio mio*.

The pitch shape of PU2 is different from the shape of PU1 in several respects. PU1 has a terminal shape and the notes are sung relatively flat (i.e. with not much vibrato). The syllable *Ah* is not greatly accentuated nor particularly loud, and it is short. In fact, it is five times shorter than the *Ah* note in PU2, even though in the score the ratio is a quarter note to a half note. The emotional state that it transmits points towards decisiveness. On the other hand, the pitch contour of PU2 goes up, involves a lot of vibrato, over-articulates *Ah* and ends at a very high frequency. In fact PU2 ends at considerably higher pitch then it begins at; something not apparent from the score. These features all imply some form of forward movement, continuation, and doubt.

VI. Conclusions

This paper presents an approach for detecting expressive patterns of the human tenor voice. We employ machine learning methods to investigate how professional opera singers manipulate sound properties such as timing, amplitude, and pitch in order to produce expressive performances of particular music fragments. We present preliminary results for performances of twelve arias by Josep Carreras. Our approach also takes into consideration features of the lyrics associated with the arias in our test suite. Currently we are considering syllable stress and type, and we are starting to look at the interplay between prosody, meter, and score, in creating lyric-dependent expressive patterns.

References

[1] Van Baelen, E. and De Raedt, L. (1996). Analysis and Prediction of Piano Performances Using Inductive Logic Programming. International Conference in Inductive Logic Programming, 55-71.

[2] H. Blockeel, L. De Raedt, and J. Ramon. Top-down induction of clustering trees. In ed. J. Shavlik, editor, Proceedings of the 15th International Conference on Machine Learning, pages 53-63, Madison, Wisconsin, USA, 1998. Morgan Kaufmann.

[3] Bresin, R. (2000). Virtual Visrtuosity: Studies in Automatic Music Performance. PhD Thesis, KTH, Sweden.

[4] Dovey, M.J. (1995). Analysis of Rachmaninoff's Piano Performances Using Inductive Logic Programming. European Conference on Machine Learning, Springer-Verlag.

[5] Friberg, A.; Bresin, R.; Fryden, L.; 2000. Music from Motion: Sound Level Envelopes of Tones Expressing Human Locomotion. Journal of New Music Research 29(3): 199-210.

[6] Gabrielsson, A. (1999). The performance of Music. In D.Deutsch (Ed.), The Psychology of Music (2nd ed.) Academic Press.

[7] Igarashi, S., Ozaki, T. and Furukawa, K. (2002). Respiration Reflecting Musical Expression: Analysis of Respiration during Musical Performance by Inductive Logic Programming. Proceedings of Second International Conference on Music and Artificial Intelligence, Springer-Verlag.

[8] Lopez de Mantaras, R. and Arcos, J.L. (2002). AI and music, from composition to expressive performance, AI Magazine, 23-3.

[9] Morales, E. (1997). PAL: A Pattern-Based First-Order Inductive System. Machine Learning, 26, 227-252.

[10] Narmour, E. (1990). The Analysis and Cognition of Basic Melodic Structures: The Implication Realization Model. University of Chicago Press.

[11] Quinlan, J.R. (1993). C4.5: Programs for Machine Learning, San Francisco, Morgan Kaufmann.

[12] Ramirez, R. et al. (2006). A Tool for Generating and Explaining Expressive Music Performances of Monophonic Jazz Melodies, International Journal on Artificial Intelligence Tools, 15(4), pp. 673-691

[13] Ramirez, R. Hazan, A. Gómez, E. Maestre, E. (2005). Understanding Expressive Transformations in Saxophone Jazz Performances, Journal of New Music Research, Vol. 34, No. 4, pp. 319-330.

[14] Rafael Ramirez, Amaury Hazan, Esteban Maestre, Xavier Serra, A Data Mining Approach to Expressive Music Performance Modeling, in Multimedia Data mining and Knowledge Discovery, Springer.

[15] Repp, B.H. (1992). Diversity and Commonality in Music Performance: an Analysis of Timing Microstructure in Schumann's 'Traumerei'. Journal of the Acoustical Society of America 104.

[16] Seashore, C.E. (ed.) (1936). Objective Analysis of Music Performance. University of Iowa Press.

[17] Serra, X. and Smith, S. (1990). "Spectral Modeling Synthesis: A Sound Analysis/Synthesis System Based on a Deterministic plus Stochastic Decomposition", Computer Music Journal, Vol. 14, No. 4.

[18] Tobudic A., Widmer G. (2003). Relational IBL in Music with a New Structural Similarity Measure, Proceedings of the International Conference on Inductive Logic Programming, Springer Verlag.

[19] Todd, N. (1992). The Dynamics of Dynamics: a Model of Musical Expression. Journal of the Acoustical Society of America 91.

[20] Widmer, G. (2002). Machine Discoveries: A Few Simple, Robust Local Expression Principles. Journal of New Music Research 31(1), 37-50.

[21] Tsur, R. (1997). Poetic Rhythm: Performance Patterns and Their Acoustic Correlates. Versification. An Electronic Journal of Literary Prosody.

Modeling Moods in Violin Performances

Alfonso Perez, Rafael Ramirez, Stefan Kersten
Universitat Pompeu Fabra
Music Technology Group
Barcelona, Spain

***Abstract*—In this paper we present a method to model and compare expressivity for different Moods in violin performances. Models are based on analysis of audio and bowing control gestures of real performances and they predict expressive scores from non expressive ones.**

Audio and control data is captured by means of a violin pickup and a 3D motion tracking system and aligned with the performed score.

We make use of machine learning techniques in order to extract expressivity rules from score-performance deviations. The induced rules conform a generative model that can transform an inexpressive score into an expressive one.

The paper is structured as follows: First, the procedure of performance data acquisition is introduced, followed by the automatic performance-score alignment method. Then the process of model induction is described, and we conclude with an evaluation based on listening test by using a sample based concatenative synthesizer.

I. Introduction

Different approaches are found in the literature for modeling expressive performances: Fryden[4] tries an analysis-by-synthesis approach, consisting of a set of proposed expressive rules that are validated by synthesis. In [3] mathematical formulae is proposed to model certain expressive ornaments. Bresin[2] and Widmer[9] make use of machine learning in order to extract expressive patterns from musical performances. In [7] they use Case Based Reasoning, that is, a database of performances that conform the knowledge of the system. In this work we follow the work done by [8], also using machine learning techniques and more specifically inductive logic programming (ILP from now on), that has the advantage of automatically finding expressive patterns without the need of an expert in musical expressivity. Regarding research in generative models, in [5] a computational model of expression in music performance is proposed.

In general this techniques try to model perceptual features such as timing deviations, dynamics or pitch. In addition, we also inform the model with control gestures, more specifically bow direction and finger position.

Apart from calculating prediction errors, models are also evaluated by listening with the help of a sampled based concatenative synthesizer under development.

Four moods are analyzed: Sadness, Happiness, Fear and Anger. Expressive features analyzed are: tempo and a set note level descriptors: onset, note duration, energy, bow direction and string being played.

In the following sections we introduce the data acquisition procedure, we detail how the model is induced and how is it performing.

II. Data Acquisition and Analysis

The training data used in our experimental investigations consist of short melodies performed by a professional violinist in four Moods: Sadness, Happiness, Fear and Anger. Pieces were played twice with and without metronome.

A set of audio and control features is extracted from the recordings and stored in a structured format. The performances are then compared to their corresponding scores in order to automatically compute the performed transformations.

The main characteristic of our data acquisition system is that of providing also motion information. This information is used for learning the model as well as for the alignment and segmentation of the performances with the scores.

A. Scores

Scores are represented as a series of notes with onset, pitch (in semitones) and duration. No extra indications are given to the performer except for the Mood. They are used to calculate performance deviations from nominal attributes of the melody.

B. Audio acquisition

Audio is captured by means of a violin bridge pickup. This way we obtain a signal not influenced by the resonances of the violin acoustic box and the room, which makes segmentation much easier than if using a microphone. From the captured audio stream we extract the audio perceptual features: frame-by-frame energy, fundamental frequency estimation and aperiodicity function. Energy is used as input for learning the model.

C. Gesture acquisition and parameter calculation

Bowing motion data is acquired by means of two 3d-motion tracker sensors, one mounted on the violin and the other on the bow as we already described in [6]. We are able to estimate with great precision and accuracy the position of the strings, the bridge and the bow. With the collected data we compute, among others, the following bowing performance parameters: bow distance to the bridge, bow transversal position, velocity and acceleration, bow force and string being played. Bow direction change and playing string are used for the segmentation and as input for learning the model.

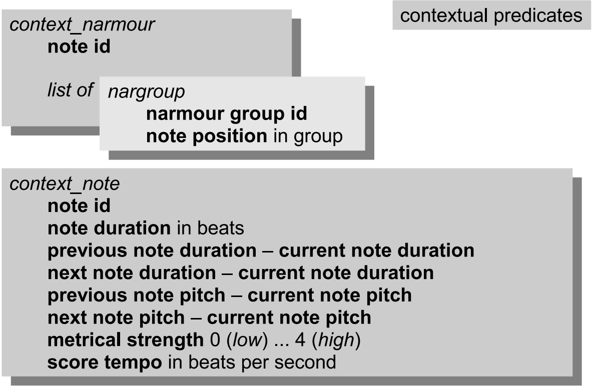

contextual predicates
context_narmour
note id
list of
nargroup
narmour group id
note position in group
context_note
note id
note duration in beats
previous note duration – current note duration
next note duration – current note duration
previous note pitch – current note pitch
next note pitch – current note pitch
metrical strength 0 (low) ... 4 (high)
score tempo in beats per second

Fig. 1. Contextual predicates.

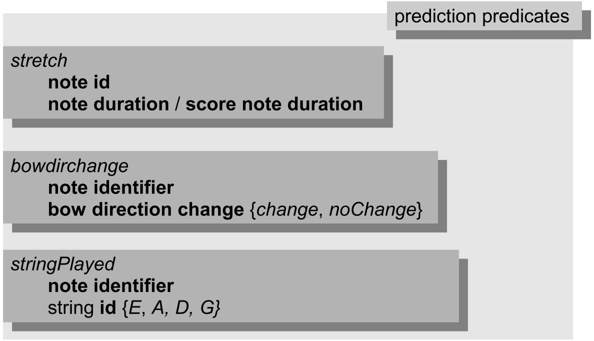


Fig. 2. Induction and Prediction predicates.

D. Score-Performance Alignment

Performances are represented with the same symbolic description as the score so that they can be aligned and deviations from the score obtained. An automatic alignment is carried out following [**?**]. It uses score information, bowing controls, and audio descriptors: A bow-direction change or a playing-string change indicates a note onset. In legato, notes segmentation is based on pitch and energy. Offsets are calculated by using energy levels. Automatic segmentation is finally manually corrected.

III. Expressive performance model

In this section we describe our inductive approach for learning the model by applying ILP techniques and we describe the evaluation results.

A. Data Description

After the alignment and segmentation, scores and expressive deviations of the performance are defined in a structured way using first order logic predicates. The musical context of each note is defined with the following predicates (Figure 1): *context_note* specifies information both about the note itself and the local context in which it appears. Information about intrinsic properties of the note includes note duration and note's metrical position, while information about its context includes the duration of previous and following notes, extension and direction of the intervals between the note and both the previous and the subsequent note, and tempo of the piece in which the note appears; *context_narmour* specifies the Narmour groups to which a particular note belongs, along with its position within a particular group. The temporal aspect of music is encoded via the predicates *pred* and *succ*. For instance, *succ(A,B,C,D)* indicates that note in position D in the excerpt indexed by the tuple (A,B) follows note C.

Expressive deviations in the performances are encoded using 4 predicates (Figure 2): *stretch* specifies the stretch factor of a given note with regard to its duration in the score; *bowdirchange* identifies points of change in bow direction; *stringPlayed* specifies in which string a note was played in the performance (certain pitches can be played in different strings resulting in a different timbre); and *dynamics* specifies the mean energy of a given note. These 4 predicates are also used for model prediction.

The use of first order logic for specifying the musical context of each note is much more convenient than using traditional attribute-value (propositional) representations. Encoding both the notion of successor notes and Narmour group membership would be cumbersome using a propositional representation. In order to mine the structured data we used Tilde's top-down decision tree induction algorithm ([1]). Tilde can be considered as a first order logic extension of the C4.5 decision tree algorithm: instead of testing attribute values at the nodes of the tree, Tilde tests logical predicates. This provides the advantages of both propositional decision trees (i.e. efficiency and pruning techniques) and the use of first order logic (i.e. increased expressiveness). The increased expressiveness of first order logic not only provides a more elegant and efficient specification of the musical context of a note, but it provides a more accurate predictive model.

B. Model Evaluation

We obtained correlation coefficients of 0.80 and 0.83 for the duration transformation and note onset prediction tasks respectively and we obtained a correctly classified instances percentage of 82% and 86% for the bow direction and string played prediction. These numbers were obtained by performing 10-fold cross-validation on the training data.

Additionally to the model performance error coefficients, listening tests are also carried as a perceptual evaluation of the models. For this we make use of a sample-based spectral concatenative synthesizer.

IV. Conclusions

We presented a model for expressive performances based not only on perceptual features but also informed with bowing. We introduced the procedure to acquire the data, learn the model and synthesize its predictions. The results seem to capture the expressive features performed. We obtained high prediction correlation coefficients and realistic synthesis of predicted performances.

V. Acknowledgments

This work was partly supported by the Spanish Ministry of Education and Science under Grant TIN2006-14932-

C02-01 (ProSeMus Project) and Yamaha Corp.

REFERENCES

[1] H. Blockeel, L. D. Raedt, and J. Ramon. Top-down induction of clustering trees. In *Proceedings of the 15th International Conference on Machine Learning*, 1998.

[2] R. Bresin. An artificial neural network model for analysis and synthesis of pianists performance styles. *JASA*, 105(2):1056, 1999.

[3] M. Clynes. *SuperConductor: The Global Music Interpretation and Performance Program*, 1998.

[4] L. Fryden, J. Sundberg, and A. Askenfelt. What tells you the player is musical? an analysis-by-synthesis study of music performance. *Publication issued by the Royal Swedish Academy of Music*, 39:61–75, 1983.

[5] P. N. Juslin, A. Friberg, and R. Bresin. Toward a computational model of expression in music performance: The germ model. *Musicae Scientiae,* Special Issue:63–122, 2002.

[6] E. Maestre, J. Bonada, M. Blaauw, A. Perez, and E. Guaus. Acquisition of violin instrumental gestures using a commercial emf device. In *Proceedings of International Computer Music Conference*, Copenhagen, Denmark, 2007.

[7] R. Mantaras, X. Serra, and J. L. Arcos. Saxex: A casebased reasoning system for generating expressive musical performances. In *Proceedings of International Computer Music Conference*, 1997.

[8] R. Ramirez, A. Hazan, E. Maestre, and X. Serra. A genetic rule-based expressive performance model for jazz saxophone. *Computer Music Journal*, 32(1):338–350, 2008.

[9] G. Widmer. Learning about musical expression via machine learning: A status report. In *17th National Conference on Artificial Intelligence*, 2000.

RetroSpat: a Perception-Based System for Semi-Automatic Diffusion of Acousmatic Music

Joan Mouba, Sylvain Marchand, Boris Mansencal, and Jean-Michel Rivet
SCRIME / LaBRI – CNRS, University of Bordeaux 1, France

***Abstract*—We present the RetroSpat system for the semi-automatic diffusion of acousmatic music. This system is intended to be a spatializer with perceptive feedback. More precisely, RetroSpat can guess the positions of physical sound sources (*e.g.* loudspeakers) from binaural inputs, and can then output multichannel signals to the loudspeakers while controlling the spatial location of virtual sound sources. Together with a realistic binaural spatialization technique taking into account both the azimuth and the distance, we propose a precise localization method which estimates the azimuth from the interaural cues and the distance from the brightness. This localization can be used by the system to adapt to the room acoustics and to the loudspeaker configuration. We propose a simplified sinusoidal model for the interaural cues, the model parameters being derived from the CIPIC HRTF database. We extend the binaural spatialization to a multi-source and multi-loudspeaker spatialization system based on a static adaptation matrix. The methods are currently implemented in a real-time free software. Musical experiments are conducted at the SCRIME, Bordeaux.**

I. Introduction

Composers of acousmatic music conduct different stages through the composition process, from sound recording (generally stereophonic) to diffusion (multiphonic). During live interpretation, they interfere decisively on spatialization and coloration of pre-recorded sonorities. For this purpose, the musicians generally use a(n un)mixing console. With two hands, this becomes hardly tractable with many sources or speakers.

The RetroSpat system supports artistically interpretation and technically room calibration. It includes a multi-source and multi-loudspeaker spatializer, that adapts to different loudspeaker configurations by "listening to the room". This involves source localization and spatialization in azimuth and distance. Here, we focus on the case of a single source with speakers in the horizontal plane.

First, we enhance the binaural model proposed by Viste [1]. We propose to simplify the spatial cues model, resulting in a new sinusoidal model with better mathematical properties and comparable errors using the CIPIC database [2]. Second, we also consider the distance of the source, with a localization based on the brightness.

Last but not least, we extend the binaural spatialization to a multi-loudspeaker spatialization system. In the classic VBAP [3] approach, the control of the interaural-level difference (ILD) is done in a frequency-independent and pair-wise way that was previously used for source panning. But this method is suitable only for frequencies up to 600Hz. The RetroSpat system also operates on loudspeakers in a pair-wise manner. But the computation of the coefficients for each channel is based on an adaptation matrix of head-related transfer functions (HRTFs), leading to complex and frequency-dependent coefficients.

This paper is organized as follows. In Section II, we present some generalities in acousmatic music and we highlight some practical weaknesses to be improved. After an extensive presentation of the model in Section III, we describe the associated spatialization and localization methods in Sections IV and V, respectively. Section VI is dedicated to the presentation of the RetroSpat software.

II. Acousmatic Music

A. History

Over centuries, the music has continuously undergone various innovations. In 1948, Schaeffer and Henry at the "Radio Télévision Française" were interested in the expressive power of sounds. They used microphones to capture sounds, discs as supports, and transformation tools. The *musique concrète* was born.

In 1949, Eimer gave birth to *electronic music* in the studios of the German radio "Nordwestdeutscher Rundfunk" in Cologne. This music was produced by frequency generators. Koenig and Stockhausen were among the first to use it.

The merge of *musique concrète* and *electronic music* gave rise to *electro-acoustic music* or *acousmatic music*. Today, many musical pieces are created worldwide. Acousmatic has become a discipline that is taught in universities and conservatories.

B. Actual Practices

Composers of acousmatic music use both electronic and natural sounds recorded close to a microphone, such as wind noise, voices, wrinkling paper, etc. The sounds are then processed by a computer and organized by editing and mixing. The result is a *musical composition*.

However, the creation gets its full value when it is played in concert using an *acousmonium*: an orchestra of loudspeakers. The acousmonium consists of a highly variable number of loudspeakers with different characteristics. The interpreter of the piece controls the acousmonium from a special (un)mixing console.

The originality of such a device is to map the two stereo channels at the entrance to 8, 16, or even hundreds of channels of projection. Each channel is controlled individually by knobs and equalization systems. The channel is assigned to one or more loudspeakers positioned according to the acoustical environment and the artistic strategy.

C. Expected Improvements

Behind his/her console, the interpreter of acousmatic music acts in real time on various sound parameters such as spatial location, sound intensity, spectral color. He/She broadcasts a unique version of the music fixed on a medium. The acousmatic diffusion requires some skills.

RetroSpat intends to facilitate the work of the interpreter by improving the following embarrassing practices:

- two wheels needed to spatialize one source;
- stereo sources as inputs;
- no individual source path, only one global mix path;
- the distance spatialization requires some expertise.

III. Binaural Model

We consider a punctual and omni-directional sound source in the horizontal plane, located by its (ρ, θ) coordinates, where ρ is the distance of the source to the head center and θ is the azimuth angle. Indeed, as a first approximation in most musical situations, both the listeners and instrumentalists are standing on the (same) ground, with no relative elevation.

The source s will reach the left (L) and right (R) ears through different acoustic paths, characterizable with a pair of filters, which spectral versions are called Head-Related Transfer Functions (HRTFs). HRTFs are frequency- and subject-dependent. The CIPIC database [2] samples different listeners and directions of arrival.

A sound source positioned to the left will reach the left ear sooner than the right one, in the same manner the right level should be lower due to wave propagation and head shadowing. Thus, the difference in amplitude or Interaural Level Difference (ILD, expressed in decibels – dB) [4] and difference in arrival time or Interaural Time Difference (ITD, expressed in seconds) [5] are the main spatial cues for the human auditory system [6].

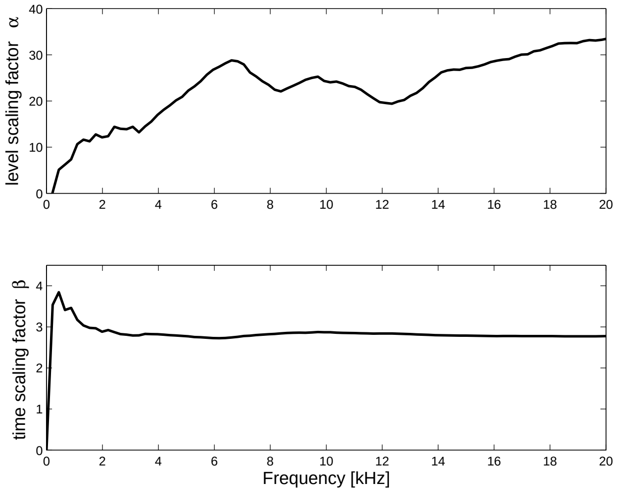


Fig. 1. *Frequency-dependent scaling factors: α (top) and β (bottom).*

A. Interaural Level Differences

After Viste [1], the ILDs can be expressed as functions of $\sin(\theta)$, thus leading to a sinusoidal model:

$$\text{ILD}(\theta, f) = \alpha(f) \sin(\theta) \tag{1}$$

where $\alpha(f)$ is the average scaling factor that best suits our model, in the least-square sense, for each listener of the CIPIC database (see Figure 1). The overall error of this model over the CIPIC database for all subjects, azimuths, and frequencies is of 4.29dB. The average model error and inter-subject variance are depicted in Figure 2.

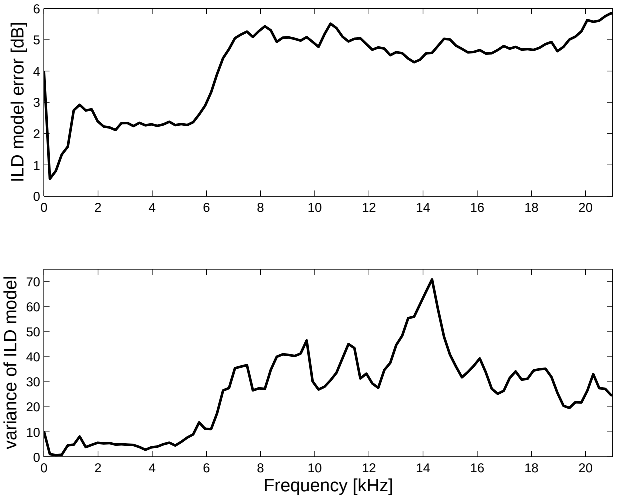


Fig. 2. *Average ILD model error (top) and inter-subject variance (bottom) over the CIPIC database.*

Moreover, given the short-time spectra of the left (X_L) and right (X_R) channels, we can measure the ILD for each time-frequency bin with:

$$\text{ILD}(t, f) \quad = \quad 20\log_{10} \left| \frac{X_L(t, f)}{X_R(t, f)} \right|. \tag{2}$$

B. Interaural Time Differences

Because of the head shadowing, Viste uses for the ITDs a model based on $\sin(\theta) + \theta$, after Woodworth [7]. However, from the theory of the diffraction of an harmonic plane wave by a sphere (the head), the ITDs should be proportional to $\sin(\theta)$. Contrary to the model by Kuhn [8], our model takes into account the inter-subject variation and the full-frequency band. The ITD model is then expressed as:

$$\text{ITD}(\theta, f) = \beta(f) r \sin(\theta)/c \tag{3}$$

where β is the average scaling factor that best suits our model, in the least-square sense, for each listener of the CIPIC database (see Figure 1), r denotes the head radius, and c is the sound celerity. The overall error of this model over the CIPIC database is 0.052ms (thus comparable to the 0.045ms error of the model by Viste). The average model error and inter-subject variance are depicted in Figure 3.

Practically, our model is easily invertible, which is suitable for sound localization, contrary to the $\sin(\theta) + \theta$ model by Viste which introduced mathematical errors at the extreme azimuths (see [9]).

Given the short-time spectra of the left (X_L) and right (X_R) channels, we can measure the ITD for each time-frequency bin with:

$$\text{ITD}_p(t, f) \quad = \quad \frac{1}{2\pi f} \left(\angle \frac{X_L(t, f)}{X_R(t, f)} + 2\pi p \right). \tag{4}$$

The coefficient p outlooks that the phase is determined up to a modulo 2π factor. In fact, the phase becomes ambiguous above 1500Hz, where the wavelength is shorter than the diameter of the head.

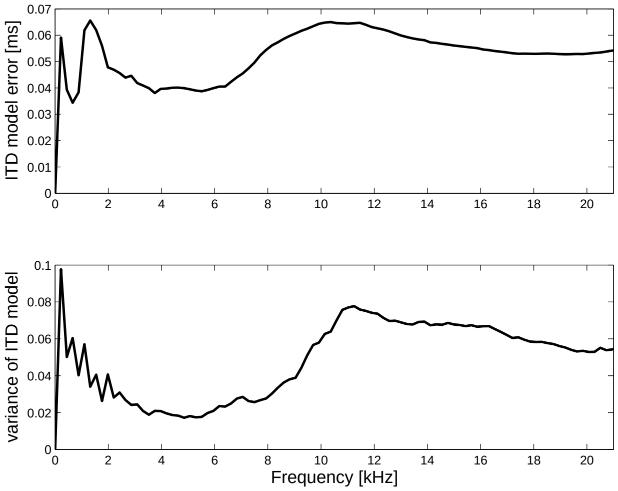


Fig. 3. *Average ITD model error (top) and inter-subject variance (bottom) over the CIPIC database.*

C. Distance Cues

The distance estimation or simulation is a complex task due to dependencies on source characteristics and the acoustical environment. Four principal cues are predominant in different situations: intensity, direct-to-reverberant (D/R) energy ratio [10], spectrum, and binaural differences (noticeable for distances less than 1m, see [11]). Their combination is still an open research subject. Here, we focus effectively on the intensity and spectral cues.

In ideal conditions, the intensity of a source is halved (decreases by -6dB) when the distance is doubled, according to the well-known Inverse Square Law [12]. Applying only this frequency-independent rule to a signal has no effect on the sound timbre. But when a source moves far from the listener, the high frequencies are more attenuated than the low frequencies. Thus the sound spectrum changes with the distance. More precisely, the spectral centroid moves towards the low frequencies as the distance increases. In [13], the authors show that the frequency-dependent attenuation due to atmospheric attenuation is roughly proportional to f^2, similarly to the ISO 9613-1 norm [14]. Here, we manipulate the magnitude spectrum to simulate the distance between the source and the listener (see Section IV). Conversely, we measure the spectral centroid (related to brightness) to estimate the source's distance to listener (see Section V).

IV. Spatialization

A. Relative Distance Effect

In a concert room, the distance is often simulated by placing the speaker near / away from the auditorium, which is sometimes physically restricted in small rooms. In fact, the architecture of the room plays an important role and can lead to severe modifications in the interpretation of the piece.

Here, simulating the distance is a matter of changing the magnitude of each short-term spectrum X. More precisely, the ISO 9613-1 norm [14] gives the frequency-dependent attenuation factor in dB for given air temperature, humidity, and pressure conditions. At distance ρ, the magnitudes of $X(f)$ should be attenuated by $D(f,\rho)$ decibels:

$$D(f,\rho) = \rho \cdot a(f). \tag{5}$$

where $a(f)$ is the frequency-dependent attenuation, which will have an impact on the brightness of the sound (higher frequencies being more attenuated than lower ones).

More precisely, the total absorption in decibels per meter $a(f)$ is given by a rather complicated formula:

$$\begin{aligned}\frac{a(f)}{P} \approx \quad & 8.68 \cdot F^2 \Big\{ 1.84 \cdot 10^{-11} \left(\tfrac{T}{T_0}\right)^{\frac{1}{2}} P_0 + \left(\tfrac{T}{T_0}\right)^{-\frac{5}{2}} \\ & \quad \Big[0.01275 \cdot e^{-2239.1/T} / [F_{r,O} + (F^2/F_{r,O})] \\ & \quad +0.1068 \cdot e^{-3352/T} / [F_{r,N} + (F^2/F_{r,N})] \Big] \Big\} \end{aligned} \tag{6}$$

where $F = f/P$, $F_{r,O} = f_{r,O}/P$, $F_{r,N} = f_{r,N}/P$ are frequencies scaled by the atmospheric pressure P, and P_0 is the reference atmospheric pressure (1 atm), f is the frequency in Hz, T is the atmospheric temperature in Kelvin (K), T_0 is the reference atmospheric temperature (293.15K), $f_{r,O}$ is the relaxation frequency of molecular oxygen, and $f_{r,N}$ is the relaxation frequency of molecular nitrogen. See [13] for details.

B. Binaural Spatialization

In binaural listening conditions using headphones, the sound from each earphone speaker is heard only by one ear. Thus the encoded spatial cues are not affected by any cross-talk signals between earphone speakers.

To spatialize a sound source to an expected azimuth θ, for each short-term spectrum X, we compute the pair of left (X_L) and right (X_R) spectra from the spatial cues corresponding to θ, using Equations (1) and (3), and:

$$X_L(t,f) = X(t,f) \cdot 10^{+\Delta_a(f)/2} e^{+j\Delta_\phi(f)/2}, \tag{7}$$
$$X_R(t,f) = X(t,f) \cdot 10^{-\Delta_a(f)/2} e^{-j\Delta_\phi(f)/2} \tag{8}$$

(because of the symmetry among the left and right ears), where Δ_a and Δ_ϕ are given by:

$$\Delta_a(f) = \text{ILD}(\theta,f)/20, \tag{9}$$
$$\Delta_\phi(f) = \text{ITD}(\theta,f) \cdot 2\pi f. \tag{10}$$

The control of both amplitude and phase should provide better audio quality [15] than amplitude-only spatialization[1] (see below).

Indeed, we reach a remarkable spatialization realism through informal listening tests with AKG K240 Studio headphones. The main problem which remains is the classic front / back confusion [16].

C. Multi-Loudspeaker Spatialization

In a stereophonic display, the sound from each loudspeaker is heard by both ears. Thus, the stereo sound is filtered by a matrix of four transfer functions ($C_{ij}(f,\theta)$) between loudspeakers and ears (see Figure 4). Here, we generate the paths artificially using the binaural model. The best panning coefficients under CIPIC conditions for the pair of speakers to match the binaural signals at the ears (see Equations (7) and (8)) are then given by:

$$K_L(t,f) = C \cdot (C_{RR}H_L - C_{LR}H_R), \tag{11}$$
$$K_R(t,f) = C \cdot (-C_{RL}H_L + C_{LL}H_R) \tag{12}$$

[1] see URL: `http://dept-info.labri.fr/˜sm/SMC08/`

with the determinant computed as:

$$C = 1/\left(C_{LL}C_{RR} - C_{RL}C_{LR}\right). \qquad (13)$$

In extreme cases where $|C| = 0$ (or close to zero) at any frequency, the matrix C is ill-conditioned, and the solution becomes unstable. To avoid unstable cases, attention should be paid during the loudspeakers configuration stage, before live diffusion.

During diffusion, the left and right signals (Y_L, Y_R) to feed left and right speakers are obtained by multiplying the short-term spectra X with K_L and K_R, respectively:

$$\begin{aligned} Y_L(t,f) &= K_L(t,f) \cdot X(t,f), & (14)\\ Y_R(t,f) &= K_R(t,f) \cdot X(t,f). & (15)\end{aligned}$$

In a setup with many speakers we use the classic pairwise paradigm [17], consisting in choosing for a given source only the two speakers closest to it (in azimuth): one at the left of the source, the other at its right.

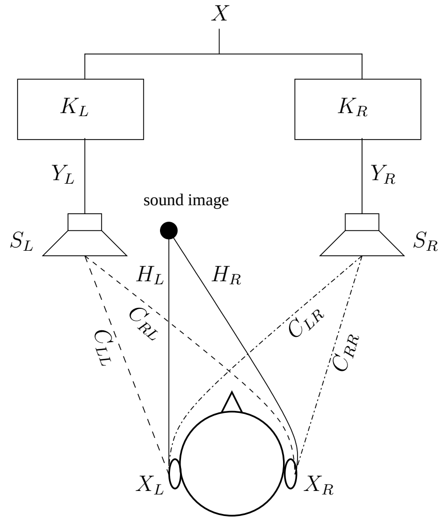


Fig. 4. *Stereophonic loudspeaker display.*

D. Analysis of Panning Coefficients

We used the speaker pair $(-30°, +30°)$ to compute the panning coefficients at any position (between the speakers) with the two techniques: our approach and the classic vector-based amplitude panning (VBAP) approach [3]. VBAP was elaborated under the assumption that the incoming sound is different only in amplitude, which holds for frequencies up to 600Hz. In fact, by controlling correctly the amplitudes of the two channels, it is possible to produce resultant phase and amplitude differences for continuous sounds that are very close to those experienced with natural sources [16]. We restrict our comparisons to the $[0, 800]$Hz frequency band.

1) Comparisons of Panning Coefficients: The panning coefficients of the two approaches are very similar until 600Hz (see Figure 5), and can differ significantly above. In fact, our coefficients are complex values, and their imaginary parts can contribute in a significant way (see Figure 6).

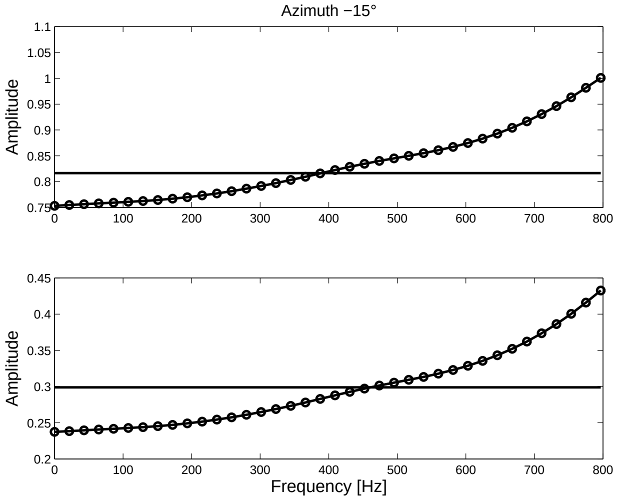


Fig. 5. *Amplitude of the panning coefficients from VBAP (plain) and our approach (dotted), for the left (top) and right (bottom) channels of the panning pair for $-15°$, in the $[0, 800]$Hz band.*

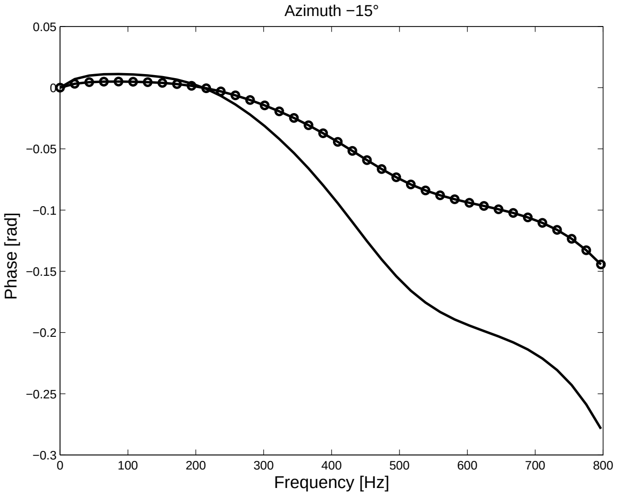


Fig. 6. *Phase of the panning coefficients from our approach, for the left (dotted) and right (plain) channels of the panning pair for $-15°$, in the $[0, 800]$Hz band.*

2) Comparisons of the Ratio of Panning Coefficients: Generally, inter-channel differences are perceptually more relevant (*e.g.* ILD, ITD) than absolute values.

Given the left and right panning coefficients, K_L and K_R, we compute the *panning level difference* (PLD):

$$\text{PLD} = 20\log_{10}\left|\frac{K_L}{K_R}\right|. \qquad (16)$$

We computed the absolute difference between the PLDs of both VBAP and our approach. The maximal PLD difference (in the considered frequency band) has a linear trend, and its maximum does not exceed 3dB. Thus, the two approaches seem to be consistent in the $[0, 800]$Hz band (see Figure 7). For higher frequencies, the new approach should yield better results, as confirmed perceptively in our preliminary and informal listening tests.

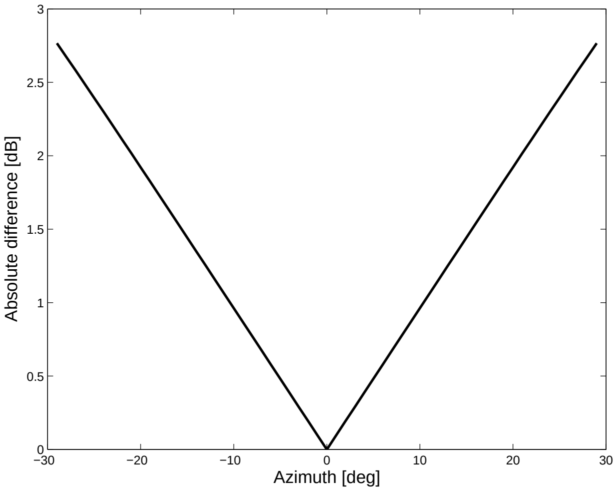


Fig. 7. *Maximum difference per azimuth between PLDs of VBAP and the proposed method in the* $[0, 800]$*Hz band.*

V. Localization

A. Azimuth Estimation

In Auditory Scene Analysis (ASA), ILDs and ITDs are the most important cues for source localization. Lord Rayleigh mentioned in his Duplex Theory [18] that the ILDs are more prominent at high frequencies (where phase ambiguities are likely to occur) whereas the ITDs are crucial at low frequencies (which are less attenuated during their propagation).

Obtaining an estimation of the azimuth based on the ILD information (see Equation (2)) is just a matter of inverting Equation (1):

$$\theta_L(t,f) = \arcsin\left(\frac{\text{ILD}(t,f)}{\alpha(f)}\right). \tag{17}$$

Similarly, using the ITD information (see Equation (4)), to obtain an estimation of the azimuth candidate for each p, we invert Equation (3):

$$\theta_{T,p}(t,f) = \arcsin\left(\frac{c \cdot \text{ITD}_p(t,f)}{r \cdot \beta(f)}\right). \tag{18}$$

The $\theta_L(t,f)$ estimates are more dispersed, but not ambiguous at any frequency, so they are exploited to find the right modulo coefficient p that unwraps the phase. Then the $\theta_{T,p}(t,f)$ that is nearest to $\theta_L(t,f)$ is validated as the final θ estimation for the considered frequency bin, since it exhibits a smaller deviation:

$$\theta(t,f) = \theta_{T,m}(t,f), \tag{19}$$

$$\text{with} \quad m = \text{argmin}_p \left|\theta_L(t,f) - \theta_{T,p}(t,f)\right|.$$

Practically, the choice of p can be limited among two values ($\lceil p_r \rceil$, $\lfloor p_r \rfloor$), where

$$p_r = \left(f \cdot \text{ITD}(\theta_L, f) - \frac{1}{2\pi}\angle\frac{X_L(t,f)}{X_R(t,f)}\right). \tag{20}$$

An estimate of the azimuth of the source can be obtained as the peak in an energy-weighted histogram (see [9]). More precisely, for each frequency bin of each discrete spectrum, an azimuth is estimated and the power corresponding to this bin is accumulated in the histogram at this azimuth. For the corresponding bin frequency f, the power $|X(f)|^2$ is estimated by inverting Equations (7) and (8) for the left and right spectra, respectively, then the square of the estimate of the loudest – supposedly most reliable – channel is retained for the power estimate.

Thus, we obtain a power histogram as shown in Figure 8. This histogram is the result of the localization of a Gaussian white noise of 0.5s spatialized at azimuth $-45°$. On this figure, we can clearly see two important local maxima (peaks), one around azimuth $-45°$, the other at azimuth $-90°$. The first (and largest) one corresponds to the sound source; the second one is a spurious peak resulting from extreme ILDs (a problem we have to solve in our future research).

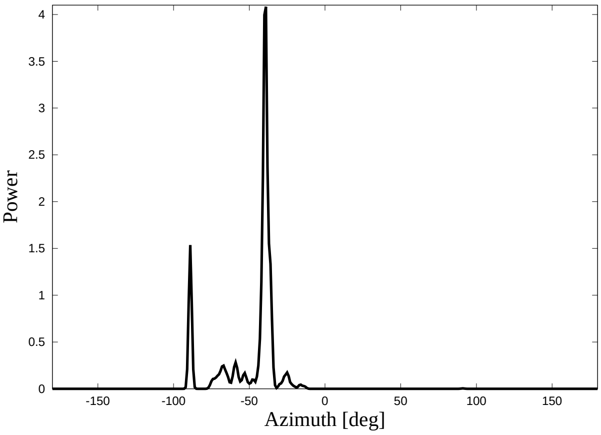


Fig. 8. *Histogram obtained with a source at azimuth* $-45°$*. One can clearly see two important local maxima (peaks): one around azimuth* $-45°$*, the other at azimuth* $-90°$*. The first (and largest) one corresponds to the sound source; the second one is a spurious peak resulting from extreme ILDs.*

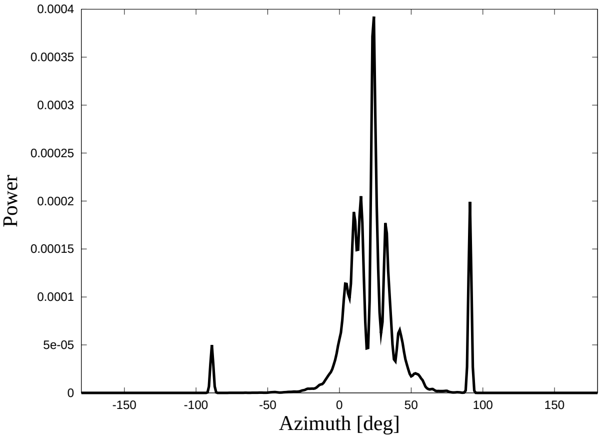


Fig. 9. *Histogram obtained with a real source positioned at azimuth* $30°$ *in a real room, with binaural signals recorded at the ears of the musician.*

For our localization tests, we spatialized a Gaussian white noise using convolutions with the HRTFs of the KEMAR manikin (see [2]), since they were not part of the database used for the learning of our model coefficients and thus should give results closer to those expected with a real – human – listener. Indeed, in our first experiments with real listeners (see Figure 9), the same trends as in Figure 8 were observed: a rather broader histogram but still with a local maximum close to the azimuth of the sound source, plus spurious maxima at extreme azimuths $\pm 90^\circ$.

To verify the precision of the estimation of the azimuth, we spatialized several noise sources at different azimuths in the horizontal plane, between -80° and $+80^\circ$, and we localized them using the proposed method. The results are shown in Figure 10. We observe that the absolute azimuth error is less than 5° in the $[-65, +65]^\circ$ range.

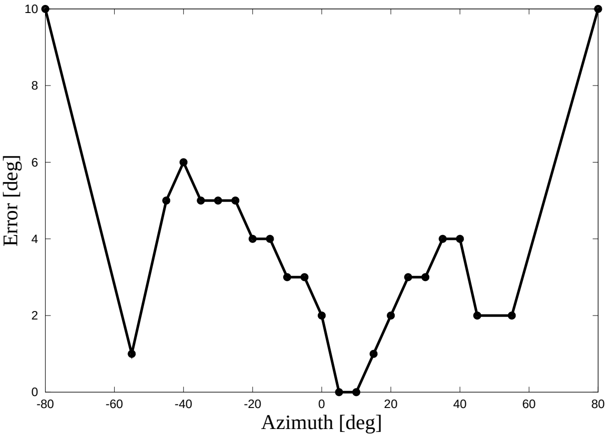


Fig. 10. *Absolute error of the localization of the azimuth from Gaussian white noise spatialized at different azimuths using convolutions with the HRTFs of the KEMAR manikin.*

In real reverberant environments, due to more superpositions at the microphones, an amplitude-based method is not really adapted; in contrast, generalized cross-correlation based ITD estimation should be more robust [19].

B. Distance Estimation

As a reference signal for distance estimation, we use a Gaussian white noise spatialized at azimuth zero, since pure tones are not suitable for distance judgments [20]. The distance estimation relies on the quantification of the spectral changes during the sound propagation in the air.

To estimate the amplitude spectrum, we first estimate the power spectral density of the noise using the Welch's method [21], [22]. More precisely, we compute the mean power of the short-term spectra over L frames, then take its square root, thus:

$$|X| = \sqrt{\frac{1}{L}\sum_{l=-(L-1)/2}^{l=+(L-1)/2} |X_l|^2}. \tag{21}$$

In our experiments, we consider $L = 21$ frames of $N = 2048$ samples, with an overlap factor of 50% (and with a CD-quality sampling rate of 44.1kHz, thus the corresponding sound segment has a length < 0.5s).

Then we use this amplitude spectrum to compute the spectral centroid:

$$\mathcal{C} = \frac{\sum_f f \cdot |X(f)|}{\sum_f |X(f)|}. \tag{22}$$

The spectral centroid moves towards low frequencies when the source moves away from the observer. The related perceptive brightness is an important distance cue.

We know the reference distance since the CIPIC speakers were positioned on a 1-m radius hoop around the listener. By inverting the logarithm of the function of Figure 11, obtained thanks to the ISO 9613-1 norm and Equations (5), (6), and (22), we can propose a function to estimate the distance from a given spectral centroid:

$$\begin{aligned} \rho(\log(\mathcal{C})) &= -38.89044\mathcal{C}^3 + 1070.33889\mathcal{C}^2 \\ &- 9898.69339\mathcal{C}^+30766.67908 \end{aligned} \tag{23}$$

given for the air at 20° Celsius temperature, 50% relative humidity, and 1 atm pressure.

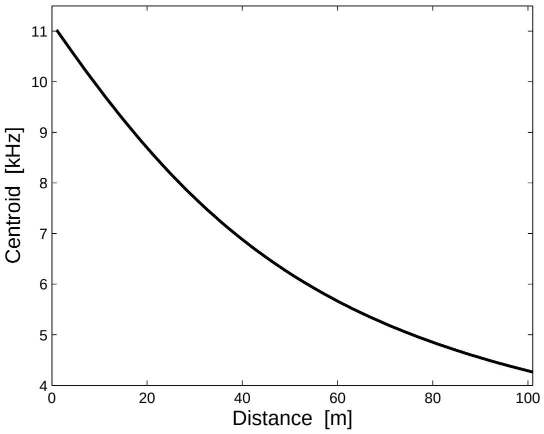


Fig. 11. *Spectral centroid (related to perceptive brightness) as a function of distance at 20° Celsius temperature, 50% relative humidity, and 1 atm atmospheric pressure (for white noise played at CD quality).*

Up to 25m, the maximum distance error is theoretically less than 4mm, if the noise power spectral density is known. However, if the amplitude spectrum has to be estimated using Equation (21), then the error is greater, though very reasonable until 50m. Figure 12 shows the results of our simulations for Gaussian white noise spatialized at different distances in the $[0, 100]$m range.

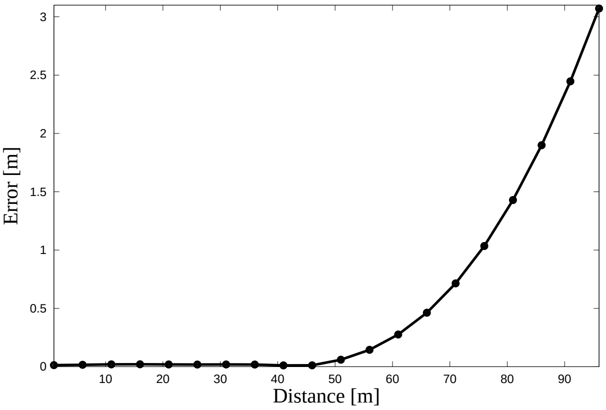


Fig. 12. *Absolute error of the localization of the distance from Gaussian white noise spatialized at different distances.*

VI. RetroSpat Software System

The RetroSpat system is being implemented as a real-time musical software under the GNU General Public

License (GPL). The actual implementation is based on C++, Qt4[2], JACK[3], FFTW[4] and works on Linux and MacOS X.

Currently, RetroSpat implements the described methods (*i.e.* localization and spatialization) in two different modules: *RetroSpat Localizer* for speaker setup detection and *RetroSpat Spatializer* for the spatialization process. We hope to merge the two functionalities in one unique software soon.

A. RetroSpat Localizer

RetroSpat Localizer (see Figure 13) is in charge of the automatic detection of the speakers configuration. It also allows the user to interactively edit a configuration, which has been just detected or loaded from an XML file.

The automatic detection of the positions (azimuth and distance) of the speakers connected to the soundcard is of great importance to adapt to new speaker setups. Indeed, it will be one of the first actions of the interpreter in a new environment.

For room calibration, the interpreter carries headphones with miniature microphones encased in earpieces (see Figure 15, where Sennheiser KE4-211-2 microphones have been inserted in standard headphones). The interpreter orients the head towards the desired zero azimuth. Then, each speaker plays in turn a Gaussian white noise sampled at 44.1kHz. The binaural signals recorded from the ears of the musician are transferred to the computer running RetroSpat Localizer. Each speaker is then localized in azimuth and distance. The suggested configuration can be adjusted or modified by the interpreter according to the rooms characteristics.

B. RetroSpat Spatializer

For sound spatialization, mono sources are loaded in RetroSpat, parameterized, and then diffused. The settings include the volume of each source, the initial localization, the choice of special trajectories such as circle, arc, etc. A loudspeaker-array configuration is the basic element for the spatialization (see Section VI-A).

The snapshot of Figure 14 depicts a 7-source mix of instruments and voices (note icons), in a 6-speaker front-facing configuration (loudspeaker icons), obtained from RetroSpat Localizer.

During the diffusion, the musician can interact individually with each source of the piece, change its parameters (azimuth and distance), or even remove / insert a source from / into the scene. In this early version, the interaction with RetroSpat is provided by a mouse controller.

Thanks to an efficient implementation using the JACK sound server, RetroSpat can diffuse properly simultaneous sources even within the same speaker pair (see Figure 14, three sources in speaker pair (2,3)). All the speaker pairs have to stay in synchrony. To avoid sound perturbation, the Qt-based user interface runs in a separated thread with less priority than the core signal processing process. We tested RetroSpat on a MacBook Pro, connected to 8 speakers, through a MOTU 828 MKLL soundcard, and were able to play several sources without problems. However, further testing is needed to assess scalability limits.

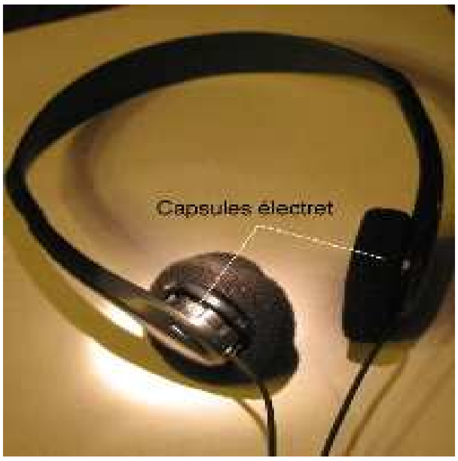


Fig. 15. *The "phonocasque" used for the binaural recordings: standard headphones where microphone capsules have been inserted.*

C. Musical Applications

In a live concert, the acousmatic musician interacts with the scene through a special (un)mixing console.

With RetroSpat, the musician has more free parameters on one single controller (mouse):

- only mouse movement to control simultaneously the azimuthal and distance location;
- mono sources as inputs;
- many sources can be spatialized to different locations at the same time;
- a dynamic visualization of the whole scene (source apparition, movement, speed, etc.) is provided.

We believe that RetroSpat should greatly simplify the interpreter interactions and thus should allow him / her to focus more on the artistic performance.

VII. Conclusions and Future Work

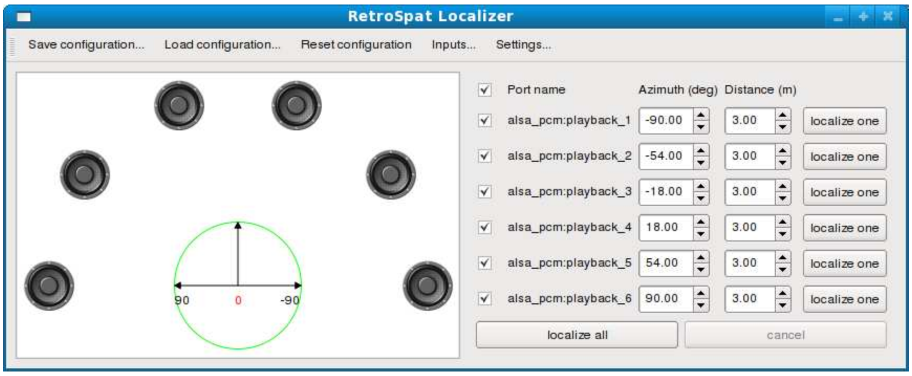
In this paper, we have introduced a flexible multi-source, multi-loudspeaker system: RetroSpat. This real-time system implements our proposed binaural to multi-loudspeaker spatialization method. The system can also locate the loudspeakers azimuths and distances.

Several experiments at the SCRIME studio on an octophonic setup justify the utility of the system for live performance by composers of electroacoustic music. Next, we should enhance the source localization in real – reverberant – environments, and possibly evolve to source control through gesture or a more intuitive hardware controller. Also, a major scientific challenge would be to separate the different sources present in a binaural mix (for a semi-automatic diffusion from a compact disc as support).

VIII. Acknowledgments

This research was carried out in the context of the SCRIME *(Studio de Création et de Recherche en Informatique et Musique Électroacoustique)* and was supported by the *Conseil Régional d'Aquitaine*, the *Ministère de la Culture*, the *Direction Régionale des Actions Culturelles d'Aquitaine*, and the *Conseil Général de la Gironde*.

[2] see URL: `http://trolltech.com/products/qt`
[3] see URL: `http://jackaudio.org`
[4] see URL: `http://www.fftw.org`



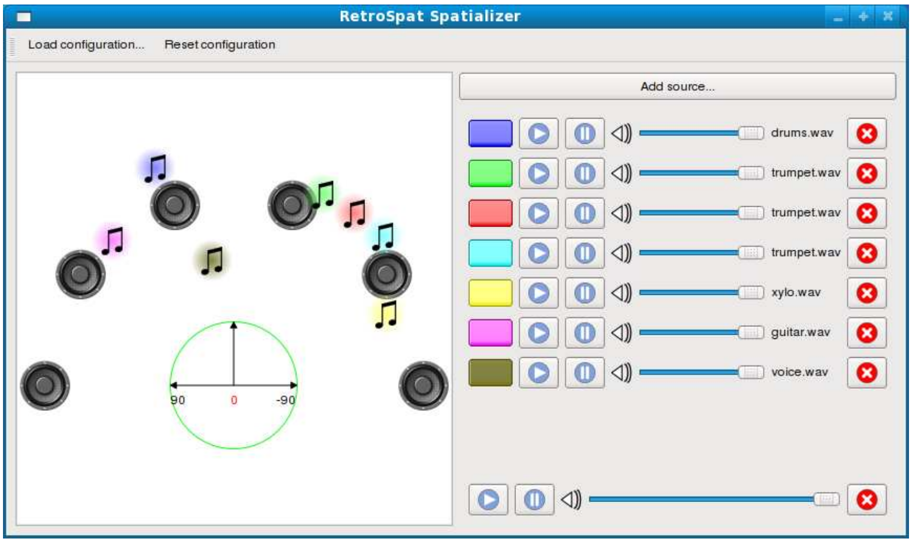
RetroSpat Localizer

Save configuration... Load configuration... Reset configuration Inputs... Settings...

☑	Port name	Azimuth (deg)	Distance (m)	
☑	alsa_pcm:playback_1	-90.00	3.00	localize one
☑	alsa_pcm:playback_2	-54.00	3.00	localize one
☑	alsa_pcm:playback_3	-18.00	3.00	localize one
☑	alsa_pcm:playback_4	18.00	3.00	localize one
☑	alsa_pcm:playback_5	54.00	3.00	localize one
☑	alsa_pcm:playback_6	90.00	3.00	localize one

localize all cancel

Fig. 13. *RetroSpat Localizer graphical user interface with a 6-speaker configuration.*



RetroSpat Spatializer

Load configuration... Reset configuration

Add source...

drums.wav
trumpet.wav
trumpet.wav
trumpet.wav
xylo.wav
guitar.wav
voice.wav

Fig. 14. *RetroSpat Spatializer graphical user interface, with 7 sources spatialized on the speaker setup presented on Figure 13.*

References

[1] H. Viste, "Binaural Localization and Separation Techniques," Ph.D. dissertation, École Polytechnique Fédérale de Lausanne, Switzerland, 2004.

[2] V. R. Algazi, R. O. Duda, D. M. Thompson, and C. Avendano, "The CIPIC HRTF Database," in *Proceedings of the IEEE Workshop on Applications of Signal Processing to Audio and Acoustics (WASPAA)*, New Paltz, New York, 2001, pp. 99–102.

[3] V. Pulkki, "Virtual Sound Source Positioning using Vector Base Amplitude Panning," *Journal of the Acoustical Society of America*, vol. 45, no. 6, pp. 456–466, 1997.

[4] J. W. Strutt (Lord Rayleigh), "On the Acoustic Shadow of a Sphere," *Philosophical Transactions of the Royal Society of London*, vol. 203A, pp. 87–97, 1904.

[5] ——, "Acoustical Observations I," *Philosophical Magazine*, vol. 3, pp. 456–457, 1877.

[6] J. Blauert, *Spatial Hearing*, revised ed. Cambridge, Massachusetts: MIT Press, 1997, translation by J. S. Allen.

[7] R. S. Woodworth, *Experimental Psychology*. New York: Holt, 1954.

[8] G. F. Kuhn, "Model for the Interaural Time Differences in the Azimuthal Plane," *Journal of the Acoustical Society of America*, vol. 62, no. 1, pp. 157–167, 1977.

[9] J. Mouba and S. Marchand, "A Source Localization / Separation / Respatialization System Based on Unsupervised Classification of Interaural Cues," in *Proceedings of the Digital Audio Effects (DAFx) Conference*, Montreal, 2006, pp. 233–238.

[10] A. W. Bronkhorst and T. Houtgast, "Auditory Distance Perception in Rooms," *Nature*, vol. 397, pp. 517–520, 1999.

[11] D. Brungart and W. Rabinowitz, "Auditory Localization of Nearby Sources," *Journal of the Acoustical Society of America*, vol. 106, pp. 1465–1479, 1999.

[12] R. E. Berg and D. G. Stork, *The Physics of Sound*, 2nd ed. Prentice Hall, 1994.

[13] H. Bass, L. Sutherland, A. Zuckerwar, D. Blackstock, and D. Hester, "Atmospheric Absorption of Sound: Further Developments," *Journal of the Acoustical Society of America*, vol. 97, no. 1, pp. 680–683, 1995.

[14] *ISO 9613-1:1993: Acoustics – Attenuation of Sound During Propagation Outdoors – Part 1: Calculation of the Absorption of Sound by the Atmosphere*, International Organization for Standardization, Geneva, Switzerland, 1993.

[15] C. Tournery and C. Faller, "Improved Time Delay Analysis/Synthesis for Parametric Stereo Audio Coding," *Journal of the Audio Engineering Society*, vol. 29, no. 5, pp. 490–498, 2006.

[16] F. Rumsey, *Spatial Audio*, 1st ed. Oxford, United Kingdom: Focal Press, 2001, reprinted 2003, 2005.

[17] J. M. Chowning, "The Simulation of Moving Sound Sources," *Journal of the Acoustical Society of America*, vol. 19, no. 1, pp. 2–6, 1971.

[18] J. W. Strutt (Lord Rayleigh), "On Our Perception of Sound Direction," *Philosophical Magazine*, vol. 13, pp. 214–302, 1907.

[19] C. H. Knapp and G. C. Carter, "The Generalized Correlation Method for the Estimation of Time Delay," *IEEE Transactions on Signal Processing*, vol. 24, no. 4, pp. 320–327, 1976.

[20] J. Molino, "Perceiving the Range of a Sound Source When the Direction is Known," *Journal of the Acoustical Society of America*, vol. 53, no. 5, pp. 1301–1304, 1973.

[21] P. D. Welch, "The Use of Fast Fourier Transform for the Estimation of Power Spectra: A Method Based on Time-Averaging over Short, Modified Periodograms," *IEEE Transactions on Audio and Electroacoustics*, vol. 15, no. 22, pp. 70–73, 1967.

[22] M. H. Hayes, *Statistical Digital Signal Processing and Modeling*. New Jersey: John Wiley & Sons, 1996.

Asymmetrical Envelope Shapes in Sound Spatialization

Francesco Canavese*, Francesco Giomi*, Damiano Meacci* and Kilian Schwoon†

* Tempo Reale, Florence, Italy, {fc, fg, dm}@centrotemporeale.it
† Hochschule für Künste, Bremen, Germany, k.schwoon@hfk-bremen.de

Abstract — **Amplitude-based sound spatialization without any further signal processing is still today a valid musical choice in certain contexts. This paper emphasizes the importance of the resulting envelope shapes on the single loudspeakers in common listening situations such as concert halls, where most listeners will find themselves in off-centre positions, as well as in other contexts such as sound installations. Various standard spatialization techniques are compared in this regard and a refinement is proposed, which results in asymmetrical envelope shapes. This method combines a strong sense of localization and a natural sense of continuity. Some examples of pratical application carried out by Tempo Reale are also discussed.**

I. Simulations and Patterns

Most contemporary research on sound spatialization focusses on the simulation of other spaces rather than the actual physical listening space. The idea of placing "an arbitrary (possibly time-varying) location within an illusory acoustic space that we hear but do not see" [1] was pioneered by John Chowning [2] and can be found nowadays, for instance, in the sophisticated "holographic" techniques of wave field synthesis [3, 4]. This concept tries to "hide" loudspeakers as much as possible from the listeners, in order to create convincing virtual sound locations.

On the other hand, composers may wish to use loudspeakers as "instruments" and create interesting spatial patterns between them. This approach might be defined as pattern-oriented as opposed to simulation-oriented. The authors have developed a spatialization system which originated in live electronic productions by the Italian composer Luciano Berio. His use of electronic spatialization seems to be a natural extension of the principles of his instrumental writing, where "identical notes or similar figures pass between groups that are similar in timbre, but separated in space" [5]. In this kind of musical context, a homogeneous sound quality and sonic presence is important. Spatial movements should be achieved by purely amplitude-based methods, without altering the signals using techniques such as delay, reverberation or filtering, which are generally involved in the simulation of spatial depth.

The problem of a more or less small privileged listening area (sweet spot), which characterizes simulation-oriented spatialization systems, is less relevant in a pattern-oriented approach, although patterns are usually also more evident from a central listening position. In any case, it is useful to consider not only the privileged central perspective, but to analyze what actually happens in off-centre listening positions, where the effective envelope shape applied by a spatialization algorithm to a single loudspeaker located close to the listener becomes perceptually significant.

II. Comparing Spatialization Methods

There are some advantages and disadvantages of common amplitude-based spatialization techniques that will be examined by comparing one of the most simple trajectories, a regular rotation on a circular octophonic loudspeaker setup.

In classical amplitude (or intensity) panning, transitions between adjacent loudspeakers are controlled by curves that provide a constant intensity [6]. This obviously creates a symmetrical envelope beginning at the peak of the previous loudspeaker and ending at the peak of the following one (Fig. 1). While this method works fine in a central listening position, and is also acceptable for slow movements in off-centre positions, it creates an undesirable effect of artificial interruption on the single loudspeakers when the movement becomes too fast.

Whereas the rising envelope shape is tolerable for the listener, the fast decay and the following zero amplitude have a rather disturbing quality. Belladonna and Vidolin noted this very early [7] and implemented a generic "offset" in their spatialization system (spAAce). Instead of returning to zero amplitude, a low offset amplitude is kept continuously on all speakers (Fig. 2). An interesting analogy can be observed in an implementation of the same trajectory using Ambisonics. In this spatialization technique, a sound field is constructed from directional and omnidirectional components of a previously encoded signal [8, 9]. Ambisonics implies modulations of amplitude and phase on each loudspeaker. Fig. 3 shows only the amplitude variations: depending on the weight of the omnidirectional component of the encoded signal, rather "blurred" envelope shapes are generated that never actually return to zero amplitude.

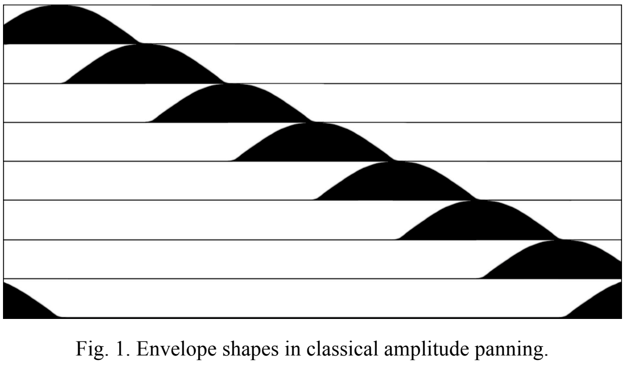


Fig. 1. Envelope shapes in classical amplitude panning.

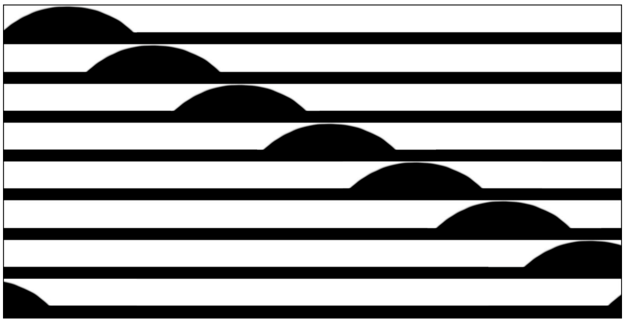


Fig. 2. Amplitude panning with offset.

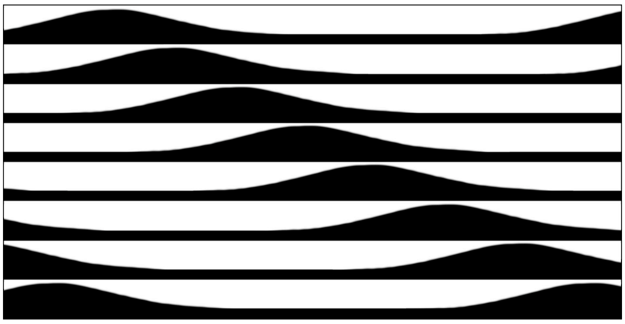


Fig. 3. Amplitude curves derived from Ambisonics.

Both spAAce and Ambisonics avoid the problem of disturbing envelope shapes at high speed on the single loudspeakers, which is typical for classical amplitude panning. But they do so by basically smoothing the movement, and therefore they lose a strong sense of localization. This is due to the fact, that in all these techniques sounds “arrive” at a certain loudspeaker in the same way they “leave” it, generating thus symmetrical envelope shapes.

III. An Asymmetrical Approach

Considering loudspeakers as instruments in a pattern-oriented approach, envelope shapes created by spatialization algorithms can be understood musically as “articulations”. In order to achieve a strong sense of localization, the sound on each loudspeaker must be rather accentuated at the beginning, whereas the decay should be relatively long, giving way smoothly to the sound on the next loudspeaker. Therefore, asymmetrical envelope shapes with a well-defined attack and a longer decay are necessary.

At Tempo Reale a set of spatialization objects for use in Max/MSP was developed [10]. These are based on linear interpolations, which in a second instance are rescaled in order to obtain constant intensity. The gain factors G for n loudspeakers are multiplied by a rescaling factor R, which is calculated as:

$$R(G) = \frac{1}{\sqrt{\sum_{i=1}^{n} G_i^2}}$$

For efficiency reasons, R is not calculated at sampling rate, but only once for each MSP signal vector (which can be reduced to a single sample in the current MSP version). Within each signal vector, the interpolation is linear. For basic transitions between two loudspeakers, this generates a light S-like curve which very gradually rises/decays near the extreme values, whereas it is relatively steep at the centre (Fig. 4). From a listening position close to a loudspeaker, this curve is often preferable to the standard square-root or sinusoidal functions used in stereo panning, which are both very steep near zero.

In the Tempo Reale spatialization system, movements are generated by scheduled sequences of lists representing gain values. The interpolation times can be defined individually for each list. Loudspeaker patterns are usually described by pseudo-binary gain values, using “1.” for the active and “0.” for the non-active speakers. If a pattern has more than one active loudspeaker at the same time, the gain factors are automatically rescaled as described above: a pseudo-binary pattern such as (1. 0. 1.) would generate the effective gain factors (0.71 0. 0.71). Actually, it is possible to choose arbitrary lists of gain factors, as they only represent proportions.

A rotation is simply generated by a sequence of lists scheduled at regular intervals (Table I). It is then possible to create asymmetrical envelope shapes by defining a decay factor for successive loudspeaker configurations, producing a sort of “shadow” of the previous configurations. For each new loudspeaker configuration in Table II, the previous gain values are multiplied by a constant factor d=0.5. This list is then superimposed on the current list by selecting the higher value at each position. Fig. 5 shows the corresponding envelope shapes. Fig. 6 illustrates the envelope shapes obtained by applying different decay factors to the list atoms and rescaling the linear interpolations as described above. In all these cases the curves start rising at the peak of the previous

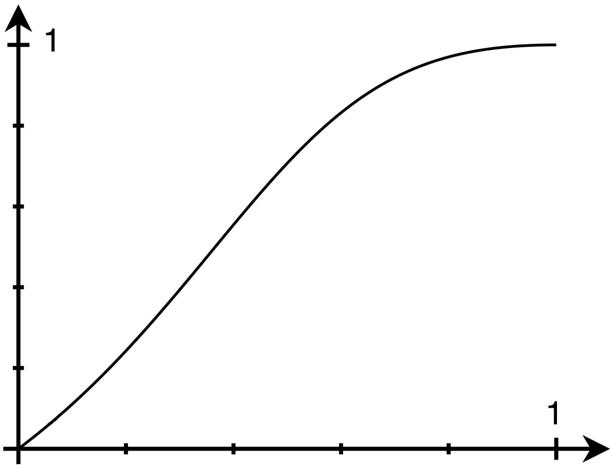


Fig. 4. S-like amplitude curve.

TABLE I.
Sequence of lists for a simple rotation.

	spk1	spk2	spk3	spk4	spk5	spk6	spk7	spk8
step1	1.	0.	0.	0.	0.	0.	0.	0.
step2	0.	1.	0.	0.	0.	0.	0.	0.
step3	0.	0.	1.	0.	0.	0.	0.	0.
step4	0.	0.	0.	1.	0.	0.	0.	0.
step5	0.	0.	0.	0.	1.	0.	0.	0.
step6	0.	0.	0.	0.	0.	1.	0.	0.
step7	0.	0.	0.	0.	0.	0.	1.	0.
step8	0.	0.	0.	0.	0.	0.	0.	1.
…								

TABLE II.
LISTS FOR A ROTATION WITH A CONSTANT DECAY FACTOR d=0.5.

	spk1	spk2	spk3	spk4	spk5	spk6	spk7	spk8
step1	1.	0.	0.	0.	0.	0.	0.	0.
step2	0.5	1.	0.	0.	0.	0.	0.	0.
step3	0.25	0.5	1.	0.	0.	0.	0.	0.
step4	0.13	0.25	0.5	1.	0.	0.	0.	0.
step5	0.06	0.13	0.25	0.5	1.	0.	0.	0.
step6	0.03	0.06	0.13	0.25	0.5	1.	0.	0.
step7	0.02	0.03	0.06	0.13	0.25	0.5	1.	0.
step8	0.01	0.02	0.03	0.06	0.13	0.25	0.5	1.
step9	1.	0.01	0.02	0.03	0.06	0.13	0.25	0.5
step10	0.5	1.	0.01	0.02	0.03	0.06	0.13	0.25
…								

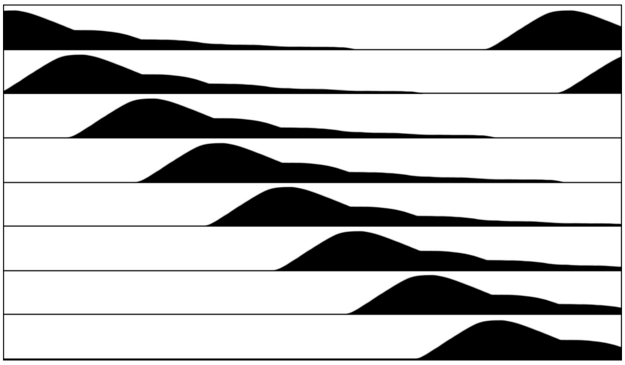


Fig. 5. Envelope shapes generated by the lists in Table II.

loudspeaker configuration, but the decay phase is more or less extended, depending on the decay factor. It can also be seen how the rising curves vary according to the gain amount distributed over the other loudspeakers.

The Tempo Reale spatialization also provides routines for generating symmetrically "blurred" gain distributions in each list, basically by applying a blur factor to adjacent loudspeakers. With a blur factor b=0.5, a list such as (0. 0. 1. 0. 0.) is, for instance, transformed into (0.25 0.5 1. 0.5 0.25). For regular rotations, this method also generates symmetrically blurred envelope shapes in time, rather similar to those of the Ambisonics example discussed above. Both methods (blurred positions and extended decays over time) can be freely combined and may create a great variety of asymmetrical shapes (Table III, Fig. 7).

IV. SOME EXAMPLES

As mentioned above, the Tempo Reale spatialization system had been initially developed for Luciano Berio's live electronic projects. His approach to spatialization in his late work was extremely pattern-oriented. He developed a notation system in which he basically defined sequences of loudspeaker configurations with holding times (t_p) and movement times (t_m) for the transitions to the next loudspeaker configurations. Fig. 8 shows the notation of a sequence with continuously changing durations of t_p and t_m. Moreover, the number of active loudspeakers in each configuration varies between one and two. Applying the usual automatic rescaling mechanisms and a decay factor, the resulting envelope shapes of such a simple sequence becomes rather complex (Fig. 9). In the current implementation, the decay factor is only applied to the scheduled lists that generate the pattern. Therefore, during the holding times there is no variation of the gain factors left over from the previous configurations (the amounts of "shadow"). This emphasizes the contrast between holding and movement, but of course other implementations, which might generate more continuous decaying envelopes, may also have musical significance. As a strategy in live electronic performance practice, the decay factors are usually decided in the preproduction phase in the studio. As it is often necessary to adapt this parameter to the specific reverberation characteristics of the actual performance space, the performance system used in these productions [11] provides efficient rescaling mechanisms for the decay factors on the level of the single sequence as well as on a global level. This allows a precise adjustment of the envelope shapes during the rehearsals.

In his live electronic projects [5], Berio only once used a classical loudspeaker setup with an octophonic circle around the audience, namely in *Ofanìm* (1988–1997) for female voice, two children's choirs, two instrumental groups, and live electronics. In his works of musical theater *Outis* (1996) and *Cronaca del Luogo* (1999) he experimented with vertical loudspeaker positions. This idea can also be found in *Altra voce* for alto flute, mezzo-soprano, and live electronics (1999), where two diverging diagonal lines of loudspeakers reach from the musicians at the center of the stage to the upper left and right corners of the concert hall. This kind of geometry only makes sense in a pattern-oriented approach to spatialization. The

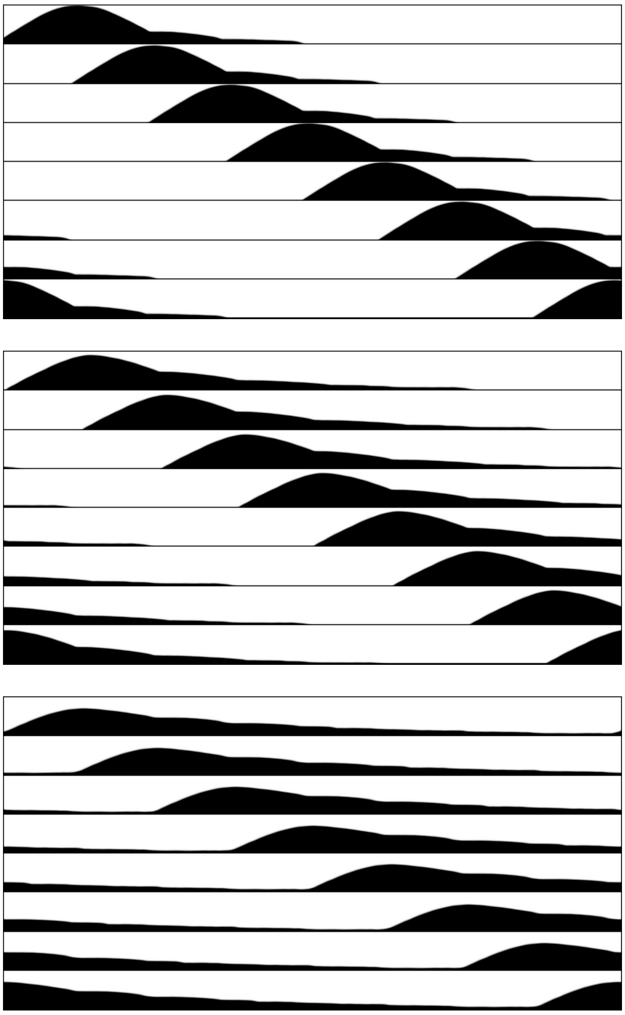


Fig. 6. Comparison of rotations with different decay factors (d=0.3/0.5/0.7).

TABLE III.
ROTATION WITH BLUR FACTOR B=0.4 AND DECAY FACTOR d=0.7.

	spk1	spk2	spk3	spk4	spk5	spk6	spk7	spk8
step1	1.	0.4	0.16	0.06	0.03	0.06	0.16	0.4
step2	0.7	1.	0.4	0.16	0.06	0.03	0.06	0.16
step3	0.49	0.7	1.	0.4	0.16	0.06	0.03	0.06
step4	0.34	0.49	0.7	1.	0.4	0.16	0.06	0.03
step5	0.24	0.34	0.49	0.7	1.	0.4	0.16	0.06
step6	0.17	0.24	0.34	0.49	0.7	1.	0.4	0.16
step7	0.16	0.17	0.24	0.34	0.49	0.7	1.	0.4
step8	0.4	0.16	0.17	0.24	0.34	0.49	0.7	1.
step9	1.	0.4	0.16	0.17	0.24	0.34	0.49	0.7
step10	0.7	1.	0.4	0.16	0.17	0.24	0.34	0.49
…								

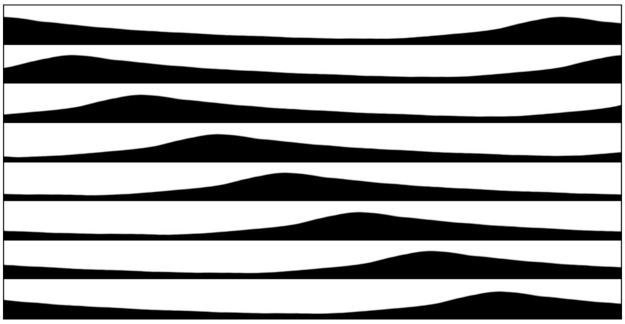


Fig. 7. Envelope shapes generated by the lists in Table III.

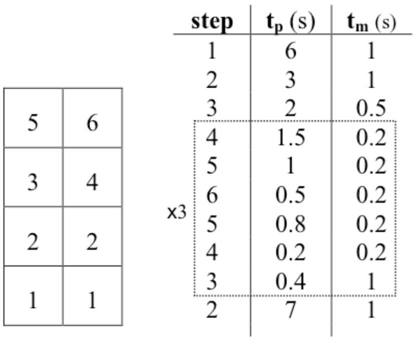
well-defined attacks of the envelope shapes are especially important in this case to make the different diagonal loudspeaker positions distinguishable to our ears.

Apart from concert productions, Tempo Reale is often also involved in projects of sound installation art, like the one realized in 2002 at the new Auditorium in Rome [12]. Only very rarely, there are central listening positions in these works: the visitors are free to move in space, in every position a valid listening experience must be possible. In most of these productions, loudspeaker distributions are chosen to create interesting relationships with particular spatial situations, emphasizing or transforming geometrical properties of an architecture. For instance, in the exhibition "Visible cities. Renzo Piano Building Workshop" (Milan, 2007), Tempo Reale realized a sound installation, entitled *Memory,* with 19 loudspeakers placed in a huge space where the central area was not accessible. The loudspeakers were hanging from the ceiling, forming different paths and listening areas (Fig. 10). Sound movements where mainly structured as linear trajectories of varying length. Even in a position directly underneath a loudspeaker, the transition to the adjacent speaker was clearly perceptible due to the clear attack of its envelope shape. At the same time the decay mechanism provided a smooth fadeout on the loudspeaker above one's head.

V. Conclusions

The Tempo Reale spatialization system is the result of practical research carried out by a team of and musicians and developers dealing with live electronic projects and sound installations. The basic effort was to overcome psychoacoustical problems of classical amplitude panning in a pattern-oriented compositional approach. Concentrating on the envelope shapes applied to the single loudspeakers signals, and transforming them with the asymmetrical methods described in this article has led to successful psychoacoustical and musical results in a great variety of situations. These strategies can be applied in a standard 5.1-channel surround sound setup, as well as with complex and non-standard loudspeakers distributions. Movements can be appreciated even in listening positions relatively close to single loudspeakers, while retaining a strong overall sense of localization.

The concept of timed sequences of loudspeaker configurations is very easy to grasp and to deal with for composers and sound artists. On the other hand, the resulting envelope shapes can be rather complex and sophisticated, having a precise control over the "articulation" of each loudspeaker by adjusting only two parameters (decay and blur factor). The flexibility of these adjustments is important, as can be understood in analogy to human interpreters: musicians more or less unconsciously change their way of playing, especially tempo, dynamics and articulation, depending on the reverberation characteristics of concert halls and their positions on stage. The possibility of easily modifying the envelope shapes of the spatialization system is a relevant new option for sound diffusion and underlines the concept of loudspeakers as "instruments". Furthermore, the very nature of a pattern-oriented approach is to avoid conflicts between a virtual space and an actual physical listening space, which may have very particular characteristics on its own.



5	6
3	4
2	2
1	1

	step	t_p (s)	t_m (s)
	1	6	1
	2	3	1
	3	2	0.5
x3	4	1.5	0.2
	5	1	0.2
	6	0.5	0.2
	5	0.8	0.2
	4	0.2	0.2
	3	0.4	1
	2	7	1

Fig. 8. Notation of a spatialization pattern by Luciano Berio.

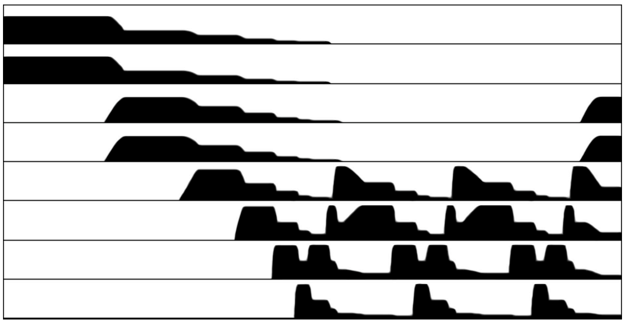


Fig. 9. Envelope shapes corresponding to Fig. 8.

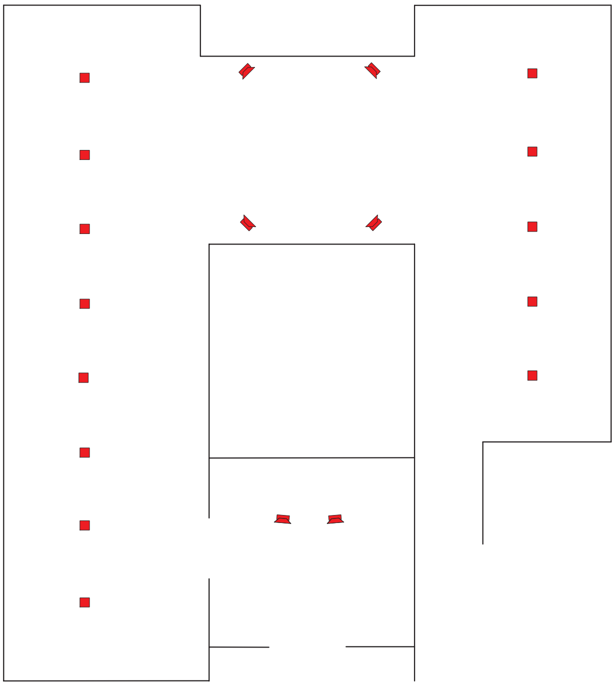


Fig. 10. Loudspeaker configuration for the installation *Memory*, realized for the exhibition “Visible cities” by Renzo Piano.

References

[1] F. R. Moore, *Elements of Computer Music*, Englewood Cliffs, New Jersey: 1990, p. 387.

[2] J. Chowning, “The simulation of moving sound sources”, *Journal of the Audio Engineering Society*, vol. 19 n. 1, pp. 2-6, 1971.

[3] A. J. Berkhout, “A holographic approach to acoustic control”, *Journal of the Audio Engineering Society*, 36(12):977-995, 1988.

[4] D. de Vries and M. M. Boone. “Wave field synthesis and analysis using array technology”, *IEEE workshop on applications of signal processing to audio and acoustics*, October 1999.

[5] F. Giomi, D. Meacci and K. Schwoon, “Live Electronics in Luciano Berio's Music”, *Computer Music Journal*, vol. 27 n. 2, 2003, pp. 31-32.

[6] see [1], pp. 353-359.

[7] A. Belladonna and A. Vidolin. “spAAce: un programma di spazializzazione per il Live Electronics”, *Proceedings of the Second International Conference on Acoustics and Musical Research*, Ferrara: 1995, pp. 113-118.

[8] M. Gerzon. “Periphony: With-Height Sound Reproduction”, *Journal of the Audio Engineering Society*, vol. 21 n. 1, pp. 2-10

[9] J. C. Schacher and P. Kocher, “Ambisonics Spatialization Tools for Max/MSP”, *Proceedings of the International Computer Music Conference*, ICMA, New Orleans: 2006.

[10] D. Zicarelli “An Extensible Real-Time Signal Processing Environment for Max”, *Proceedings of the International Computer Music Conference*, ICMA, Ann Arbor: 1998, pp. 463-466

[11] F. Canavese, F. Giomi, D. Meacci, and K. Schwoon, “An SQL-Based Control System for Live Electronics”, *Proceedings of the International Computer Music Conference*, Barcelona: 2005, pp. 753-756.

[12] F. Giomi, D. Meacci and K. Schwoon, “Electroacoustic Music in a Multi-Perspective Architectural Context: A sound installation for Renzo Piano's Auditorium in Rome”, *Organised Sound*, vol. 8 n. 2, 2003, pp. 157-162.

Distance Encoding in Ambisonics Using Three Angular Coordinates

Rui Penha
INET-MD / University of Aveiro, Portugal, ruipenha@ua.pt

***Abstract* — In this paper, the author describes a system for encoding distance in an Ambisonics soundfield. This system allows the postponing of the application of cues for the perception of distance to the decoding stage, where they can be adapted to the characteristics of a specific space and sound system. Additionally, this system can be used creatively, opening some new paths for the use of space as a compositional factor.**

I. Introduction

Sound spatialization has been one of the main aspects of interest for composers since the beginning of electroacoustic music experiments. Several techniques for spatialization were created specifically for synthesizing virtual soundfields, whilst others were derived from sound recording techniques. Amongst the latter, the Ambisonic system [1] has resurfaced in the last years, due to an increasing interest in its characteristics [2] and its advantages over other surround systems [3].

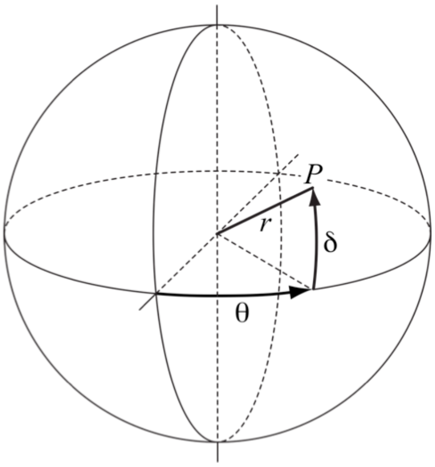


Fig. 1. Coordinate system used in this paper.

The basic first order Ambisonic system is known as the B-format, in which a full three-dimensional soundfield is decomposed in spherical harmonics and encoded into four channels, known as *W*, *X*, *Y* and *Z*. Using spherical polar coordinates - azimuth θ and elevation δ, as shown in Fig. 1 -, one can encode a signal *i* at point *P*, in a first order Ambisonics B-format, using simple equations:

$$W = i\frac{1}{\sqrt{2}},$$

$$X = i\cos\theta\cos\delta,$$

$$Y = i\sin\theta\cos\delta,$$

$$Z = i\sin\delta,$$

using the standard weighting of the 0th order *W* channel [4]. This soundfield can then be reconstructed with a variety of speaker arrays [5], including, with additional processing, binaural stereo [6], using the decoding equation for the speaker signal *s*:

$$s = \frac{1}{S}(\frac{W}{\sqrt{2}} + X\cos\theta\cos\delta + Y\sin\theta\cos\delta + Z\sin\delta),$$

where *S* represents the total number of speakers and both the azimuth θ and elevation δ represent the angles of the speaker position on the surface of the sphere defined by the concentric speaker array.

Several higher order Ambisonic systems have been developed since the original proposal of the system [4][7], augmenting the spatial resolution of the Ambisonics encoding by increasing the order of the spherical harmonics used to encode the soundfield.

II. Distance Encoding

A. State of the Art

By relying solely on two angular coordinates - azimuth θ and elevation δ - for the encoding, the traditional Ambisonic system only encodes a two-dimensional spherical coordinate system, represented by the surface of the sphere were the sounds are projected. This system, created initially for recording rather than the encoding of synthesized sources, therefore favors the encoding of the localization of sound, excluding the distance cues that are not present or permanently added to the signal at the time of encoding, as it has been proposed with the inclusion of near field compensation at the encoding stage [8].

Additionally, a sound at the exact center of the coordinate system, hence with no azimuth nor elevation and consequently absent from all the spherical coordinates except for the 0th order *W*, cannot be easily encoded with a traditional Ambisonics approach, as addressed with the creation of W-Panning [9] for the spatialization of sounds enclosing the listener.

For higher order Ambisonics, a new encoding system [8] effectively addresses the loudspeaker near field effect by compensating for it at the encoding stage, thus allowing the reproduction of sources both inside and outside the surface of the sphere defined by the speaker array. Furthermore, if the diameter of a given speaker array is different from the reference one, compensating filters can be applied prior to the decoding stage, thus enabling the diffusion of the same encoded soundfield using different speaker setups.

B. Proposed Solution

The proposed solution to encode distance in an Ambisonic-based system consists in encoding the distance *r* as a new angular coordinate, using the hyperspherical coordinates of a 3-sphere [10], effectively turning the radial coordinate *r* in the angular coordinate ρ. By varying the angle ρ between 0 (at the center of the sphere) and $\pi/2$ (at the surface of the sphere) and adding one audio channel *D*, one can create an extended B-format with distance encoding for signal *i* at point *P* using the equations:

$$W = i\frac{1}{\sqrt{2}},$$

$$X = i\cos\theta\cos\delta\sin\rho,$$

$$Y = i\sin\theta\cos\delta\sin\rho,$$

$$Z = i\sin\delta\sin\rho,$$

$$D = i\cos\rho.$$

This soundfield with distance encoding can then be decoded using the equation:

$$\begin{aligned} s = \frac{1}{S}(\frac{W}{\sqrt{2}} + X\cos\theta\cos\delta\sin\rho + Y\sin\theta\cos\delta\sin\rho \\ +Z\sin\delta\sin\rho + D\cos\rho) \end{aligned}.$$

As a result, if a signal is encoded with the maximum distance, thus at the surface of the sphere assumed to enclose the soundfield, the *D* channel is silent and all the others work as in their original form. Conversely, if a signal is encoded with distance $\rho = 0$, thus at the origin of the coordinate system, the *D* channel receives the signal with its full amplitude and all the other channels are silent, therefore loosing all the localization cues, as the signal is at the listener's position. All the versatility of traditional Ambisonics is retained, as the *W, X, Y* and *Z* channels are exactly the same as they would be on a regular system, as long as all the signals are encoded with distance $\rho = \pi/2$. As the *D* channel only encodes distance and not direction, the traditional matrices for rotation (around the *z*-axis), tilt (around the *x*-axis) and tumble (around the *y*-axis) [4] can still be used with the *X*, *Y* and *Z* channels alone.

Although the diameter of the sphere assumed at the encoding stage must be known when reconstructing the soundfield, its value can now be different than the one of the final speaker array. It is important to note, however, that one needs to decode the central position to either a real or a virtual omnidirectional speaker, the latter being then spread by the real speakers in a weighted manner. Failing to do so would cause a progressive loss of the signals encoded towards the center of the coordinate system, in a similar way as one looses the signals encoded with an elevation angle δ of $\pm\pi/2$ in a horizontal-only speaker array.

C. Distance Cues

Besides allowing the playback of strict distances between sounds in differently sized sound systems, the encoding of the distance between virtual sound sources and the center of the coordinate system can postpone the application of cues for the perception of distance - such as the loudness, atmospheric absorption and reverberation - to the decoding stage, as long as one knows the sphere size assumed during the encoding of the soundfield. This opens the door to the fine-tuning of these cues to each sound projection space from the same encoded signals.

A simple method for decoding, e.g., an extended B-format with distance encoding for a speaker array with a smaller diameter than the one of the assumed sphere would be: to decode the signals for a virtual speaker array, with the same number of speakers as the real one but with the same diameter as the assumed sphere; to decode the signal for a virtual omnidirectional speaker at the center of the assumed sphere; to apply the required distance cues to all the decoded signals, compensating for the difference in diameters between the virtual and the real speaker arrays; to play the compensated signal of each virtual speaker in the concentric array using the real speaker with the same angular location and the weighted signal of the virtual center speaker through all the real speakers.

The nature of the angular coordinate system encoding by itself caters for the smooth fades between the loudness and atmospheric absorption filters applied to the virtual center signal and to the virtual speakers located on the surface of the assumed sphere. Preliminary listening tests have shown that this robust system is capable of very convincing results, if not physically accurate ones, which, although unlikely, remains to be tested.

III. Spatialization Vocabulary

Regardless of the fact that Ambisonics excels at physically reconstructing a recorded soundfield [2][3][7][8], both traditional and new vocabulary for spatialization can be implemented using its standard techniques. The rotation, tilt and tumble matrices, used to alter the microphone position within a recorded B-Format, can be used to continuously rotate a soundfield for creative spatialization proposes. As an example, if one implements some kind of time-dependent processing with a feedback loop, as a reverb or a delay, to an encoded soundfield and rotates the resulting soundfield while rotating the original in the opposite direction, a moving tail is created. A single parameter - the angular rotation velocity, responsible for both the direction and the size of the moving tail - can be used to manipulate the resulting effect in real-time.

By inserting stages of decoding and re-encoding around effects affecting only specific spaces, one can create small spaces within the soundfield where sounds are transformed solely by "traveling" through them. An implementation of this (in higher order Ambisonics) using Max/MSP is shown in Fig. 2. Again, the nature of the angular coordinate system encoding by itself caters for the smooth fades between different areas.

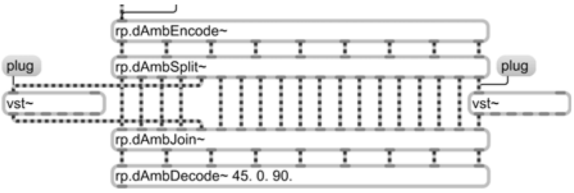


Fig. 2. Applying effects to small areas of the encoded soundfield.

The “rp.dAmbJoin~” object shown in Fig. 2 is in fact nothing more than a group of 19 encoders for specific positions within the assumed sphere: elevations $\delta = \pi/2$ and $\delta = -\pi/2$, 16 equally spaced azimuths, all with distance $\rho = \pi/2$, and the center with distance $\rho = 0$. This object can therefore be used to directly encode signals from fragmented sources, e.g. spectral spatialization [12] or streams of grains in granular synthesis.

The above approaches, albeit profiting from the distance encoding, are nevertheless possible to implement using the existing Ambisonic system. Some specific creative approaches, on the other hand, arise from the described proposal. As an example, the diameter of the assumed sphere is employed as a constant while decoding a soundfield for different speaker arrays. If used as variable, however, one can manipulate the perceived size of the diffusion space and the relative distance of all the sound objects encoded in a soundfield in real-time, e.g. by varying the size of a rotating soundfield proportionally to its angular velocity. This effect, while certainly possible to implement using existing techniques, would involve the simultaneous manipulation of numerous effects and parameters, as opposed to solely two variables, when using the proposed solution.

IV. Future Work

A. Higher Order Ambisonics

This system can naturally be expanded to higher order Ambisonics, therefore allowing greater spatial resolution. The proposed standard for the system being developed uses a stream of nine audio channels with a mixed-order system of 3rd horizontal order and 1st vertical order, thus reflecting the ubiquity of horizontal-only speaker arrays while preserving the full-sphere capabilities, with the lower resolution justified by the poorer precision of localization in the median plane [11].

B. Compatibility

The viability of higher order Ambisonics with distance encoding is conditioned by the research of a way to compensate for the near field effect of speakers [8] at the decoding stage.

The relative size of sound sources is also a very important parameter when working with distance encoding. As in the vast majority of sound spatialization techniques, the proposed system encodes sounds as dimensionless points in space. This will be addressed by researching ways to integrate the O-format [9], an Ambisonics B-format that encodes a pattern of sound radiation of an object, within the proposed system.

C. Develpment of Modular Tools for Spatialization

Modular tools for the field application of the proposed system are being created, both as Max/MSP abstractions and externals and as standalone applications. The modular approach for the creation of these tools will allow the scalability of systems within a consistent approach. At the same time, some examples of the specific vocabulary for interacting with the system are being composed. While trying to integrate some of the current Ambisonics-related research into the proposed system, the focus on the creation of straightforward, yet powerful, means of interacting with the tools under development will most certainly require a proposal for a new, modular environment for electroacoustic composition where spatialization plays the chief role.

V. Conclusion

An ongoing research, the proposed Ambisonics system effectively addresses the need to integrate distance cues in the spatialization of sound without being tied to a specific speaker array.

Acknowledgment

The author would like to thank his supervisor, João Pedro Oliveira, for his commitment to this project and inspirational work as a composer, as well as the University of Aveiro for supporting his work.

References

[1] M. Gerzon, “Periphony: with height sound reproduction,” *Journal of the Audio Engineering Society,* vol 21:1, pp. 2-10, January/February 1973.

[2] J. Bamford, *An Analysis of Ambisonic Sound Systems of First and Second Order,* unpublished MSc thesis, 1996.

[3] D. Malham, “Homogeneous and nonhomogeneous surround sound systems,” *AES UK “Second Century of Audio” Conference,* London, UK, June 1999.

[4] D. Malham, *Higher order Ambisonic systems,* http://www.york.ac.uk/inst/mustech/3d_audio/higher_order_ambisonics.pdf, retrieved in April 2008.

[5] F. Hollerweger, *Periphonic Sound Spatialization in Multi-User Virtual Environments,* Santa Barbara: CREATE, 2006.

[6] M. Noisternig, A. Sontacchi, T. Musil and R. Höldrich, “A 3D ambisonic based binaural sound reproduction system,” *AES 24th International Conference,* Banff, Canada, June 2003.

[7] J. Daniel and S. Moreau, “Further study of sound field coding with higher order ambisonics,” *AES 116th Convention,* Berlin, Germany, May 2004.

[8] J. Daniel, “Spatial sound encoding including near field effect: introducing distance coding filters and a viable, new ambisonic format,” *AES 23rd International Conference,* Copenhagen, Denmark, May 2003.

[9] D. Menzies, “W-panning and O-format, tools for object spatialization,” *AES 22nd International Conference of Virtual, Synthetic and Entertainment Audio,* Espoo, Finland, June 2002.

[10] Wikipedia, “3-Sphere,” http://en.wikipedia.org/wiki/3-sphere, retrieved in April 2008.

[11] G. Kendall, “A 3-D sound primer: directional hearing and stereo reproduction,” *Computer Music Journal,* vol 19:4, pp. 23-46, Winter 1995.

[12] R. Torchia and C. Lippe, “Techniques for Multi-Channel Real-Time Spatial Distribution Using Frequency-Domain Processing,” *Proceedings of the 2004 Conference on New Interfaces for Musical Expression,* Hamamatso, Japan, June 2004.

Granular Sound Spatialization Using Dictionary-Based Methods

Aaron McLeran*, Curtis Roads*, Bob L. Sturm†, John J. Shynk†
*Media Arts and Technology Program, University of California, Santa Barbara, CA 93106-6065, USA
†Department of Electrical and Computer Engineering, University of California, Santa Barbara, CA 93106-9560, USA

***Abstract*—We present methods for spatializing sound using representations created by dictionary-based methods (DBMs). DBMs have been explored primarily in applications for signal processing and communications, but they can also be viewed as the analytical counterpart to granular synthesis. A DBM determines how to synthesize a given sound from any collection of grains, called *atoms*, specified in a *dictionary*. Such a granular representation can then be used to perform spatialization of sound in complex ways. To facilitate experimentation with this technique, we have created an application for providing real-time synthesis, visualization, and control using representations found via DBMs. After providing a brief overview of DBMs, we present algorithms for spatializing granular representations, as well as our application program *Scatter*, and discuss future work.**

I. Introduction

Sound can be spatialized on multiple time scales [1]. In classic electronic music, many compositions are characterized by a global spatial perspective, such as a uniform blanket of reverberation applied to the entire macrostructure of a composition, e.g., Oskar Sala's *Elektronische Impressionen* (1978). In other works, spatial variations articulate mesostructural boundaries: phrases and sections. For example, Stockhausen's *Kontakte* (1960) contrasted sounds in foreground/background relationships on a time scale of phrases within moments [2].

Later, through the development of music programming languages and digital audio editors, the time scales of spatial transformations were reduced down to the level of individual sound objects. A cascading sequence of sound objects, each emanating from a different virtual space, provides the dimension of spatial depth to an otherwise flat perspective and articulates a varying topography.

Below the level of individual sound objects is the world of microsound [1]. Gabor proposed that all sound could be decomposed into a family of functions obtained by time and frequency shifts of acoustic "quanta" [3], [4]. The composer Xenakis extended Gabor's theory and proposed its inverse: any given sound can be composed, or synthesized, by elementary sonic *grains* [5]. Today, it is possible to decompose and recompose sound by a variety of means. Some methods, such as granulation, work directly in the signal time domain [1]. However, in the dictionary-based methods (DBMs) described later in this paper, a granular representation of a signal is provided through time-frequency analysis. By means of these techniques, spatialization can now be explored down to the microsound level of sonic structure, where individual spatial positions are assigned to every sonic grain.

II. Dictionary-Based Methods

DBMs provide an alternative to time-frequency signal representations, such as those made by short-term Fourier and wavelet analyses. While Fourier analysis is built upon complex sinusoids, and wavelet analysis uses the dilation of a mother wavelet, DBMs allow for any set of functions – collectively called the *dictionary*. The general idea behind DBMs is to avoid making an a priori decision about a basis that best represents a particular signal; instead, the representation basis is allowed to adapt to the signal statistics [6]. This can result in representations that are more sparse, efficient, and meaningful than those found by standard analysis methods [6], [7]. So far, DBMs have been primarily applied in applications of communications and signal processing (e.g., see [8]–[10]). Research using DBMs for sound transformation applications has only recently begun [11]–[13].

In DBMs, a signal is represented as a linear combination of waveforms chosen from a predefined dictionary of possible waveforms. Let the signal be denoted by the K-dimensional column vector $\mathbf{x}$, and let the dictionary be denoted by the matrix $\mathbf{D}_{K\times N}$, where each column is an individual waveform. The signal $\mathbf{x}$ can thus be written as

$$\mathbf{x} = \mathbf{D}\mathbf{s} \tag{1}$$

where $\mathbf{s}$ is a column vector of N weights. Observe that if $\mathbf{D}^H$ is the complex conjugate transpose of the orthonormal discrete Fourier transform matrix, then $\mathbf{s}$ is simply the Fourier transform of $\mathbf{x}$. In general, however, $N \gg K$ and $\mathbf{D}$ is overcomplete, meaning that rank$(\mathbf{D}) = K$. This implies that there will always exist at least one solution $\mathbf{s}$ satisfying (1), and possibly an infinite number of solutions. In general, without specifying any constraints, finding a solution to (1) is an ill-posed problem. Constraining the solution $\mathbf{s}$ to have the minimum number of nonzero elements creates an NP-hard problem [14]. A more relaxed constraint involves minimizing the ℓ_1-norm of $\mathbf{s}$, which creates a convex problem solvable by a linear program [7]. An entirely different set of methods for solving (1) are based on gradient descent [6], [15].

A. Matching Pursuit Algorithm

The matching pursuit (MP) algorithm is quite simple, and fast implementations exist [16]. MP iteratively builds the representation basis by choosing atoms from a given dictionary $\mathbf{D} = [\mathbf{d}_1|\mathbf{d}_2|\cdots|\mathbf{d}_N]$, where each column $\mathbf{d}_i$ is a unique waveform. At step $n+1$, a column is selected from $\mathbf{D}$ that has the largest magnitude inner product with the nth-order residual signal

$$\mathbf{g}_n = \arg\max_{\mathbf{d}\in\mathbf{D}} \big|\mathbf{d}^T\mathbf{r}(n)\big|/||\mathbf{d}|| \qquad (2)$$

where $\mathbf{r}(n) = \mathbf{x} - \widetilde{\mathbf{x}}(n)$ ($\mathbf{r}(0) \equiv \mathbf{x}$), and $\widetilde{\mathbf{x}}(n)$ is the nth-order approximation waveform ($\widetilde{\mathbf{x}}(0) \equiv \mathbf{0}$). The complexity of finding each atom in MP is on the order of computing a fast Fourier transform of the entire signal [6], [16]. After choosing $\mathbf{g}_n$, its corresponding weight is computed as

$$a_n = \mathbf{g}_n^T\mathbf{r}(n)/||\mathbf{g}_n||. \qquad (3)$$

The $(n+1)$st-order residual signal is then given by

$$\mathbf{r}(n+1) = \mathbf{r}(n) - a_n\mathbf{g}_n, \qquad (4)$$

and the algorithm repeats until some stopping criterion is met. After n iterations, the nth-order approximation of the original signal $\mathbf{x}$ is given by:

$$\widetilde{\mathbf{x}}(n) = [\mathbf{g}_0|\mathbf{g}_1|\cdots|\mathbf{g}_{n-1}]\begin{bmatrix} a_0 \\ a_1 \\ \vdots \\ a_{n-1}\end{bmatrix} \triangleq \mathbf{G}(n)\mathbf{a}(n). \qquad (5)$$

If the dictionary is at least complete, i.e., $\text{rank}(\mathbf{D}) = K$, then $\widetilde{\mathbf{x}}(n)$ will converge to the original signal $\mathbf{x}$ [6]. While MP does not guarantee that this will occur after a finite number of steps, orthogonal MP does [15] – but at a higher computational cost. For our applications, however, the approximations created by MP provide a useful and meaningful representation of the original signal.

B. Building and Specifying Dictionaries

Dictionaries are often constructed from a combination of discretized, scaled, translated, and modulated lowpass functions. For instance, a dictionary element can be parametrically described by

$$g(k) = h(k-u;s)\cos\big([k-u]\omega(k-u) + \phi(k-u)\big) \qquad (6)$$

where $0 \le k \le K-1$ is a time index, $0 \le u < K - s/2$ is a translation, $1 \le s \le K$ is a scale in samples, and $0 \le \omega(k) \le \pi$ and $0 \le \phi(k) < 2\pi$ are the modulation frequency and phase, respectively, which might depend on time, such as chirps [17]. The function $h(k;s)$ can be likened to a window function. For a Gabor atom [3], [6], $h(k;s)$ is the Gaussian function

$$h(k;s) = \begin{cases} \exp\big(-\frac{(k-s/2)^2}{2(\alpha s)^2}\big), & k = 0,1,\ldots,s-1 \\ 0, & \text{else}\end{cases} \qquad (7)$$

where α controls the variance, and s is the scale. A plot of an example Gabor atom is shown in Fig. 1.

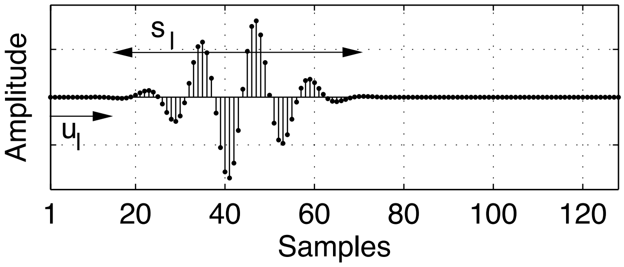


Fig. 1. Example Gabor atom with scale s_l and translation u_l.

A dictionary is created by combining numerous atoms with various scales, translations, and modulation frequencies. In contrast to Fourier and wavelet transforms, this produces a dictionary which tiles the time-frequency plane in multiple ways [18].

III. Spatialization Using Dictionary-Based Methods

After the MP algorithm is performed to a satisfactory signal-to-residual ratio (SRR), the results of the decomposition, i.e., the chosen atoms and weights, are stored as a collection of indices from the dictionary in what is called a *book*. Because each atom is parameterized, many unique sound transformations are possible [11]–[13]. This paper presents recent experiments using novel atom spatialization techniques and a variety of basis functions such as Gabor atoms, or damped sinusoids.

A. Two-Dimensional Spatialization Coordinate

In order to simplify our initial experimentation, spatialization was restricted to a circular two-dimensional (2D) array of m channels. A general 2D spatial coordinate is specified by the parameter $p \in [0,1]$. For the case of a stereo channel ($m = 2$), p is the stereo panning parameter. In a more general case, p is interpolated across the m channels such that each channel contains a number which represents the amount of an atom in that channel.

Assuming that the 2D array of m channels is circular, p wraps around to remain within its specified range. For example, $p = 1.1$ wraps to $p = 0.1$. In this paper, we assume that p is a singular spatial point and not a spatial distribution.

B. Random Scattering

If the atoms are scattered by an amount $\sigma \in [0,1]$, the spatial coordinate p for each atom becomes simply

$$p = \sigma. \qquad (8)$$

If σ is a number generated from a uniform distribution and is unique for each atom, the result is maximum spatial scattering, because each atom occupies a unique position in space. If every atom is spatialized by the same σ, either randomly generated or manually set, then the result is the opposite of scattering: instead, the entire book is localized to a singular spatial position.

C. Blur

In order to achieve spatial blur, another spatial parameter is added to σ in (8), yielding

$$p = \sigma + \beta \tag{9}$$

where β is a number generated from any desired probability distribution supported on the bounded interval $[-r, r]$. If σ is the same for all atoms in a book and β is uniquely generated for each atom, then the result is a spatial blur localized at σ.

D. Convergence and Divergence

If in (9) the interval range of β is reduced to zero ($r \to 0$), the spatialization will simply become (8). If this occurs over some time interval, and σ is the same for every atom of a book, the effect is spatial convergence to the spatial location specified by σ. On the other hand, if $r \to x$ where $x \in [0, 1]$, the result is spatial divergence.

E. Panning

Panning is distinct from scattering, convergence, and divergence, in that sound appears to move dynamically through space. Because atoms are typically of a very short duration, dynamically changing p for each atom has no perceivable effect. Therefore, the illusion of panning is achieved by individually spatializing atoms such that each atom's spatial coordinate p is set according to a global function $f(u)$ where u is the atom's time translation from (6). Thus, we can write

$$p = f(u) \tag{10}$$

where $f(u)$ is defined according to any desired process. For example, it might be a slowly varying low-frequency oscillator (LFO), a manually defined break-point function set from a graphical user interface (GUI), or some other algorithmic or stochastic process.

F. Spatializing According to Parametric Filtering

All atomic parameters from (6) are available for the construction of unique spatializing algorithms. For example, because transients are typically composed of very short duration atoms, the following rule spatially moves transient atoms of a book differently than tonal atoms:

$$p = \begin{cases} f(u), & s < \alpha \\ g(u), & \text{else} \end{cases} \tag{11}$$

where s is an atom's scale (duration) value, and α is a tunable threshold below which atoms are most likely part of a transient structure. $f(u)$ and $g(u)$ are different functions which depend on an atom's translation parameter; they can be defined according to any desired procedure. The result is the spatial dislocation of a sound's noisy transients and its harmonic tonals. Many such algorithms for spatial scattering or spatial motion are possible via a desired combination or filtering of atomic parameters.

G. Stochastic Panning

Setting σ and β from (9) to be stochastic functions that depend on atomic translation (similar to $f(u)$ in (10)) leads to fully dynamic and stochastic spatialization techniques. For example, multiple clusters of atoms, built from filtering the book according to any number of desired atomic parameters, might expand or contract into spatial clouds which move across a spatial field at unique varying rates.

IV. Scatter: A Real-Time Application Program for Manipulating Atomic Representations

A. Implementation Details

The software for Scatter was written in C++ and Objective-C using Mac OS X's Cocoa API. The synthesis was performed using a software toolkit currently under development in the Media Arts and Technology (MAT) Program at UCSB. The implementation assumes traditional block processing of groups of samples at a fixed rate, and follows well-known techniques for real-time granular synthesis [19]. However, instead of scheduling atoms within a block of samples according to purely synthetic procedures, they are scheduled according to their temporal location within a time-sorted decomposition book derived via the MP algorithm.

For a time-sorted dictionary, atom scheduling is usually not problematic as long as the dictionary is queried only for the sample range of the currently executing block as opposed to the entire book, which may contain many thousands of atoms. However, if the oscillators used for atoms are sine waves, scheduling issues may arise when the atom density as a function of time at any point in the book is extreme. Large atom densities typically correspond to complex components within a given signal, such as transients or noise. Therefore, instead of using computed sine waves, atoms are mostly synthesized using a simple sine oscillator based on a two-pole resonator, which requires only one multiply and add per sample computation. The downside to using resonators is that they are expensive if their frequency or phase is changed. Since atoms are typically of very short duration, little benefit is achieved when individual atomic parameters are changed within the block, so this is usually an acceptable compromise.

However, DBMs allow any type of waveform to be included in the dictionary, and it is possible that atoms may have durations on the order of seconds or longer. Long-duration atoms need to have the ability to change their parameters in real-time in order to avoid unwanted artifacts. Therefore, long-duration atoms are synthesized using a relatively simple computed third-order polynomial sine wave that can dynamically change its parameters with essentially no increase in the computation time.

B. Visualizing Decompositions

In order to accurately represent the energy content of individual atoms, they are represented graphically using their Wigner-Ville distribution (WVD) [20]. The WVD of

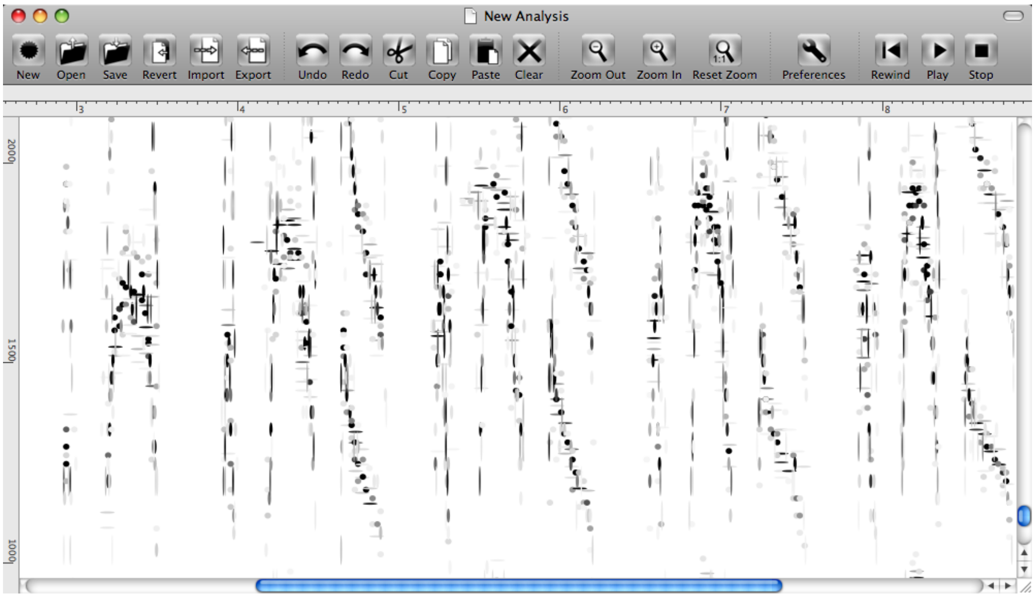


Fig. 2. GUI prototype for Scatter showing WVD plot of a decomposition.

a Gabor atom is a two-dimensional Gaussian waveform centered on a modulation frequency and time translation. Figure 2 is a screenshot of an early prototype of the main GUI for Scatter, which was influenced by SPEAR [21]; the figure illustrates the WVD plot for a decomposition. We call the superposition of WVDs of the atoms in a decomposition a *wivigram*, which has proven to be useful as a means of visualizing and interacting with atomic decompositions.

C. GUI Components

Selection, Filtering, Parametric Transformations: Several options are available for selecting individual atoms within the decomposition. They can be chosen individually, via lasso or box, or by using bounding regions in frequency and time. Once selected, the atoms can be transformed according to any of the atomic parameters such as time and frequency translation, compression, or dilation. Atoms may also be deleted, copied, or pasted.

GUI for Spatialization: A set of GUI controls has been designed which allow the user to specifically dictate the various techniques mentioned in this paper for controlling the stochastic spatial parameters. The GUI uses standard controls such as sliders, knobs, and break-point functions, which when combined with any of the selection and editing controls shown in Fig. 2, allow a user to apply any of the previously mentioned spatialization algorithms.

D. *Extensions to* Scatter

Molecular Selection and Transformation: Currently, the implementation allows only selection and transformation at the atomic level. Because books consist of many thousands of atoms, it is often difficult to perform transformations on meaningful structures in a signal. For example, it is currently difficult to select and transform individual harmonics. Thus, current work is focusing on the development of algorithms which automatically construct higher level molecular models of the decomposition and allow for intuitive GUI control and manipulations of molecules. However, these techniques are still experimental and have not yet been implemented for Scatter.

Analysis Stage: Real-time synthesis is currently being implemented using books analyzed from the Matching Pursuit Toolkit (MPTK) [16]. This has allowed development efforts to focus on real-time synthesis and GUI interactions and processing rather than the MP implementation. However, in order to fully take advantage of the unique benefits of DBMs, Scatter should include access to the analysis, and allow users to easily customize dictionaries, or set analysis parameters such as SRR to define a desired model order.

V. Future Work: Micropluriphony in the Allosphere

Stereophony, quadraphony, and octophony refer to sound positioning in a symmetrical lateral array in front of or around the listener. Periphony extends this scheme to the vertical dimension [22]. Using techniques such as wave field synthesis, the notion of periphony is extended to pluriphony: the projection of three-dimensional (3D) sounds from a variety of positions above, below, and within the audience.

MAT's current testbed for spatialization is the Allosphere at UCSB [23]. The Allosphere is a three-story-high spherical instrument in which virtual environments and performances can be experienced with full 360-degree immersion. The space is now being equipped with high-resolution active stereo projectors, a 3D sound system with several hundred speakers, and with tracking and interaction mechanisms.

Our work on spatializing atomic decompositions using DBMs has so far been focused on 2D spatialization techniques. However, current efforts are underway to extend the spatialization methods discussed here to full 3D spatialization within the Allosphere. There are many technical challenges, particularly those of scale. As is common to granular synthesis in general, spatialization of atomic decompositions faces an explosion in the number of parameters that are needed for the control of the position and movement of possibly thousands of sound events per second. A similar problem of scale arises when projection is extended from a 2D spatial field to a fully pluriphonic space with potentially hundreds of channels, such as in the Allosphere.

ACKNOWLEDGMENT

The authors would like to thank MAT Ph.D. students Lance Putnam and Graham Wakefield for consultations on various implementation details of Scatter. This work was supported in part by the National Science Foundation under Grant CCF-0729229.

REFERENCES

[1] C. Roads, *Microsound*, MIT Press, Cambridge, MA, 2001.

[2] C. Roads, "Decyphering Stockhausen's *Kontakte*," in preparation 2008.

[3] D. Gabor, "Theory of communication," *J. IEE*, vol. 93, no. 3, pp. 429–457, Nov. 1946.

[4] D. Gabor, "Acoustical quanta and the theory of hearing," *Nature*, vol. 159, no. 4044, pp. 591–594, May 1947.

[5] I. Xenakis, "Elements of stochastic music," *Gravensaner Blätter*, vol. 18, pp. 84–105, 1960.

[6] S. Mallat and Z. Zhang, "Matching pursuits with time-frequency dictionaries," *IEEE Trans. Signal Process.*, vol. 41, no. 12, pp. 3397–3415, Dec. 1993.

[7] S. S. Chen, D. L. Donoho, and M. A. Saunders, "Atomic decomposition by basis pursuit," *SIAM J. Sci. Comput.*, vol. 20, no. 1, pp. 33–61, Aug. 1998.

[8] R. M. Figueras i Ventura, P. Vandergheynst, and P. Frossard, "Low-rate and flexible image coding with redundant representations," *IEEE Trans. Image Process.*, vol. 15, no. 3, pp. 726–739, Mar. 2006.

[9] M. Elad and M. Aharon, "Image denoising via sparse and redundant representations over learned dictionaries," *IEEE Trans. Image Process.*, vol. 15, no. 12, pp. 3736–3745, Dec. 2006.

[10] S. Lesage, S. Krstulovic, and R. Gribonval, "Underdetermined source separation: Comparison of two approaches based on sparse decompositions," in *Proc. Int. Conf. Independent Component Analysis Blind Source Separation*, Charleston, SC, Mar. 2006, pp. 633–640.

[11] G. Kling and C. Roads, "Audio analysis, visualization, and transformation with the matching pursuit algorithm," in *Proc. Int. Conf. Digital Audio Effects*, Naples, Italy, Oct. 2004, pp. 33–37.

[12] B. L. Sturm, L. Daudet, and C. Roads, "Pitch-shifting audio signals using sparse atomic approximations," in *Proc. ACM Workshop Audio Music Comput. Multimedia*, Santa Barbara, CA, Oct. 2006, pp. 45–52.

[13] B. L. Sturm, C. Roads, A. McLeran, and J. J. Shynk, "Analysis, visualization, and transformation of audio signals using overcomplete methods," in *Proc. Int. Computer Music Conf.*, Belfast, Ireland, Aug. 2008.

[14] G. Davis, S. Mallat, and M. Avellaneda, "Adaptive greedy approximations," *J. Constr. Approx.*, vol. 13, no. 1, pp. 57–98, Jan. 1997.

[15] Y. Pati, R. Rezaiifar, and P. Krishnaprasad, "Orthogonal matching pursuit: Recursive function approximation with applications to wavelet decomposition," in *Proc. Asilomar Conf. Signals, Syst., Comput.*, Pacific Grove, CA, Nov. 1993, vol. 1, pp. 40–44.

[16] S. Krstulovic and R. Gribonval, "MPTK: Matching pursuit made tractable," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Process.*, Toulouse, France, Apr. 2006, vol. 3, pp. 496–499.

[17] R. Gribonval, "Partially greedy algorithms," in *Trends in Approximation Theory*, K. Kopotun, T. Lyche, and M. Neamtu, Eds., pp. 143–148. Vanderbilt University Press, May 2001.

[18] S. Mallat, *A Wavelet Tour of Signal Processing*, Academic Press, San Diego, CA, 2nd edition, 1999.

[19] B. Truax, "Real-time granular synthesis with a digital signal processor," *Computer Music Journal*, vol. 12, no. 2, pp. 14–26, 1988.

[20] D. Preis and V. C. Georgopoulos, "Wigner distribution representation and analysis of audio signals: An illustrated tutorial review," *J. Audio Eng. Soc.*, vol. 47, no. 12, pp. 1043–1053, Dec. 1999.

[21] M. Klingbeil, "Software for spectral analysis, editing, and synthesis," in *Proc. Int. Computer Music Conf.*, Barcelona, Spain, Sep. 2005.

[22] M. Gerzon, "Periphony: With height sound reproduction," *J. Audio Eng. Soc.*, vol. 21, no. 1, pp. 2–10, 1973.

[23] X. Amatriain, J. Castellanos, T. Höllerer, J. Kuchera-Morin, S. T. Pope, G. Wakefield, and W. Wolcott, "Experiencing audio and music in a fully immersive environment," in *Lecture Notes in Computer Science: Sense to Sound*, K. K. Jensen, R. Kronland-Martinet, and S. Ystad, Eds. Springer Verlag, Berlin, Germany, in press 2008.

VOCABULARY OF SPACE PERCEPTION IN ELECTROACOUSTIC MUSICS COMPOSED OR SPATIALISED IN PENTAPHONY

Bertrand Merlier

Université Lumière Lyon 2
Département Musique / Faculté LESLA
18, quai Claude Bernard
69365 LYON CEDEX 07
Bertrand.Merlier@univ-lyon2.fr

GETEME
(**G**roupe d'**Ét**ude sur l'**E**space
dans les **M**usiques **É**lectroacoustiques)
http://geteme.free.fr
geteme@free.fr

ABSTRACT

This paper begins with a brief introduction of the GETEME (**G**roupe d'**Ét**ude sur l'**E**space dans les **M**usiques **É**lectroacoustiques - Working Group about Space in Electroacoustic Musics), followed by an overview on its past, present and future activities. A first major achievement was the completion and publication of the "vocabulary of space electroacoustic musics…", coupled with the realization of a taxonomy of space.

Beyond this collection and clarification of these words in general use, it appears necessary to begin to connect words and sound.

The goal of our present research is to clarify or elaborate a vocabulary (a set of specialized words) allowing to describe space perception in electroacoustic (multiphonic) musics. The issue is delicate as it deals with psychoacoustics… as well as creators' or listeners' imagination.

In order to conduct this study, it was necessary to develop a battery of tests, procedures and listening collection of words describing listening space, and then counting and sorting words.

The sound descriptions quickly overlap, the words coincide with the same listening situations. A consensus seemed to emerge, revealing: 5 types of spatiality and 2 types of mobility, as well as a variety of adjectives to describe or characterize spatiality or mobility.

Keywords: taxonomy, terminology, describing space, spatial perception, musicology of space.

1. INTRODUCTION: THE INVENTION OF SPACE MUSICOLOGY!

In 2000, Thélème Contemporary[17] published a first CD of electroacoustic music composed or spatialised in DTS 5.1 [25]. Probably the first realization of this kind (in France)!

Four years later, the publication of a second CD in 5.1 DTS is again considered. Ten French composers are contacted. Eight of them answered positively to the proposal of Thélème Contemporain. This second CD in 5.1 DTS came out in the fall of 2004 [26].

Beyond several aesthetic and technical innovations, a great step is done. The fixing of these electroacoustic works also leads to the fixing of their space discourses on the media. It is now possible to listen five or ten times to the same work, in order to understand how the composer has put his music in space; to listen five or ten times the works of various composers (Francois Bayle, Jean-Marc Duchenne, Jean-Claude Risset…), in order to compare space discourses strategies.

In short, the fixing of spatialised electroacoustic works on a consumer multichannel support opens the way for space musicology! i.e. to analyze a space discourse or compare two spatialisation methods becomes possible!

Before considering the formalization or the conceptualizing of spatialisation strategies, before talking about space writing or space discourses, a first step appears to be: describing heard space phenomena. To do this, one needs a listening vocabulary: connecting words and space perception!

2. THE GETEME

The GETEME (**G**roupe d'**É**tude sur l'**E**space dans les **M**usiques **É**lectroacoustiques = Working Group about Space in Electroacoustic Musics) was founded in late 2003 by Jean-Marc Duchenne, Bertrand Merlier and Hélène Planel (see http://geteme.free.fr). It is supported by Thélème Contemporain (Association for Creation and Distribution of Computer Music, http://tc2.free.fr). In 2004-2006, it was granted by AFIM (Association Française pour l'Informatique Musicale = French Association for Promotion of Computer Music).

The main objectives of this working group are:

- to locate and identify the actors involved or concerned by these activities: creators, acousticians, psycho-acousticians, computer specialists, musicologists…;
- to realize a state of the art of knowledge and techniques;
- to clarify vocabulary and practices.

Seven or eight articles were published between 2004 and 2007 in several newspapers or international conferences (see references [1], [6], [8], [9], [10]). A website has been opened to introduce the GETEME activities and publish the research results, in addition to the founding members Web sites of the GETEME (see [14], [15], [16]).

Finally, a first book was published in November 2006: "The vocabulary of space and spatialisation in electroacoustic music", published by Delatour France[7] editor.

Other projects are under way, such as a spatialised sound examples DVD or a second book about space in electro-acoustic music, in a more didactic and literary way.

3. "THE VOCABULARY OF SPACE…"

3.1. Content and objectives

This glossary is a research work on the vocabulary in use in terms of electroacoustic musics spatialisation or sound space. It includes 390 words and 1200 definition, in about 220 pages.

The main object of this study is the **music produced or reproduced through loudspeakers**, without any kind of constraint or musical aesthetic.

This glossary has been mainly carried out thanks to a study and a compilation of words in use in terms of electro-acoustic music spatialisation in a large amount of paper or Internet publications.

The identification and analysis of the vocabulary in use by the community are expected to trig reflections about terminology and facilitate communication and exchanges between the various actors in these artistic or technical worlds.

3.2. Taxonomy of space

This word collection allowed to get a complete overview of the topic, and so to propose a systemic classification of space activities and means. This taxonomy allows to detect omissions or sense ambiguities (other than by empirical or intuitive means), as well as to explore more reliable multiple meanings of words.

The establishment of this taxonomy is presented in detail in the introduction of the "vocabulary of space". The interest and the use of this taxonomy were first presented at SMC06 ([8] in French), then a second time – in front of a completely different audience – at the EMS 06 conferences (Electroacoustic Music Studies) in Beijing (from 23 to 28 October 06) ([6] in English).

4. CONNECTING WORDS AND SOUND…

4.1. Two approaches: the composing vocabulary or the space perception vocabulary

This collection of words in use was a first step. The second step consists into refining this particular vocabulary and connecting "words" and "sounds".

The words included in the "Vocabulary of space…" clearly required sound illustrations or sound connections. Just as sound illustrations will certainly require the introduction of new "words" in order to characterize the "making" or the "hearing", le « faire » ou « l'entendre » (to quote Pierre Schaeffer's words).

Two approaches are possible, depending on whether one considers the point of view of the emitter (composer) or the one of the receiver (the auditor).

This paper investigates the question of space perception description only in the domain of pentaphonic electroacoustic musics.

The following paragraphs describe the testing process, the choice of sound examples, and then the analysis and sorting of words.

5. LISTENING TESTS OF PENTAPHONIC COMPOSED OR SPATIALISED MUSICS

5.1. Process description

Listening takes place in a room of average size (50 to 100 m2) audio neutral, equipped with a stereo 5.1 likely to read CD encoded with DTS 5.1. A group of a dozen people sitting around is at the centre of the room.

The collection of vocabulary characterizing listening to the "composed space" takes place in the following manner:

a) listening to a small music excerpt (of about one minute) on a 5.1 sound system;
b) individual thinking (not influenced), the result of which is imperatively written on paper by the auditors;
c) reading of written notes;
d) collective discussion, trial and search for clarification and possible consensus (not obligatory: differences may subsists);
e) possible re-hearing of the extract;
f) possible comments or words refining;
g) next example.

At the end of the test, the notes written by the auditors, describing each sample, are collected. These written individual notes (step b) guarantee individual reflection and stable information over time, and avoids collective influences.

The collective discussion and re-listenning process (step c, d, e and f) allow to improve vocabulary precision, as well as to write a brief synthesis note. It is also an opportunity for a didactic action: description of unknown psychoacoustics phenomena, as well as new words (or concepts) learning.

5.2. Sound examples choice

It has already been explained that our tests dealt with space perception of 5.1 musical compositions and not on acoustic space in general.

Sound examples were first selected among the two DTS 5.1 CD published by Thélème Contemporain in 2000 and 2004. Other CDs or DVDs were used to expand our choices to other aesthetic and technological processes:

- *Reverse* by French electro-jazz duet Orti & Sense [27] (this double CD offers the choice between stereo or DTS 5.1 versions);
- a demo DVD edited by the DTS company itself, including the group Eagles in a live concert, performing the famous tune: *Hotel California* [28].
- several examples made by the Swedish national radio and downloadable online: advertising jingles, audio reports, recordings of orchestral pieces in pentaphony [18].

A wide range of music spatialisation processes are used: multiphonic composition, spatialisation of stereo sources through hardware or software, reduction of an octophonic work on 5 channels, pentaphonic recording of instrumental performance, instrumental duet or trio put in space on 5 channels, and so on.

The "spatialisation strategy" criteria does not take part of the selection of works (or at least not directly). The musical works were essentially selected for their different perceptual effects. Table 1 presents this example list.

In this first step, we only looked at perception differences, without trying to characterize them.

Titre	repérage exact	CD
Bayle : *Arc, pour Gérard Grisey*	idx 1 ≥ 1'40	[25]
Bouttier : *Pianosphère*	idx 2 à 0'00	[25]
Duchenne : *D'après une brèche*	à 0'00	[25]
Merlier : *Ourania* (mvt.1)	idx 7	[25]
Favre : *Soufre noir*	à 0'00	[25]
Risset : *Resonant SoundScapes*	idx 10 début	[25]
Risset : *Resonant SoundScapes*	idx 12	[25]
Orti / Sens : *Ne pas arrêter - never*	idx 2 en stéréo idx 2 en 5.1	[27]
Merlier : *Fragulos*	idx 1 ou 6	inédit
Swedish Radio : Jingle de pub	idx 9	[18]
Mendelsohn ou Strauss	idx 10 ou 11	[18]
Swedish Radio : *Histoire sonore*	idx 12	[18]
Merlier : *Les chevaux de Ladoga*		[26]
Merlier : *Sillage*		[26]
Eagles : *Hotel California*	menu idx 5 ≥ 1'20	[28]

Table 1 : list of sound exemples

Excerpts lasts between 30 and 60 seconds.
Note: exact references of works and CD (references in brackets in column 3) are given at the end of this paper.

Excerpt title	Words written by the listeners		Collective synthesis written par B. Merlier
Bayle *Arc, pour Gérard Grisey*	on est au milieu de quelquechose, on est dans la soupière et ça bouge…	→	bain sonore / immersion
	bain sonore (avec quelques sons ponctuels), partout	→	/ ambiophonie
	espace spécifique aux timbres utilisés		
	ping-pong rapides, accélérés et ralentis	→	trilles d'espace,
	petites choses précises, granuleux, mobilité		scintillement
	chatoyant, dense, mouvant, envoûtant, flottant, irradiant		
	de la profondeur		
	on oublie les haut-parleurs		
Bouttier *Pianosphère*	espace clos sans jeu de profondeur		
	le son se déplace à la surface des membranes		
	la musique se déplace autour de nous	→	rotations, trajectoires
	mouvements prévisibles ou évidents	→	figures d'espace
	matériaux influençant le mouvement		
	toujours en mouvement, espace géométrique		
	mouvement circulaire d'un seul son à la fois, points qui tournent	→	lointain
	son qui part et qui arrive à destination, voyage		
	prise de conscience des haut-parleurs		
Duchenne *D'après une brèche*	profondeur / événements sonores distribués dans l'espace		
	les événements ne sont pas dans le même espace	→	polyphonie d'espaces
	il y a du proche et du lointain, verticalité (on perçoit l'élévation)		
	vraie composition spatiale, multidimensionnel		
	superposition d'espaces, strates, grande diversité d'effets		réverbération
	précis, clair et cinématographique		
	grands espaces, circule partout		
	parfois trajectoires, mais pas trop, joystick	→	mouvements / figures
	espaces dynamiques, du statique et du narratif		
	plans sonores timbraux, plans d'espace dynamique	→	images d'espace ou
	paysages d'événements sonores, images d'espace		phonographies
	phonographie, narratif		
Merlier *Ourania mvt1*	triangulation / à l'envers de l'habitude : plan proche au fond et plan lointain en face		création de mouvement par fragmentation
	événements ponctuels / travail par points / du vide entre les points / des endroits inhabités		
	déplacements imprévisibles, événements improbables, disparates, surprise, ping-pong, réponses		discontinuité
	espaces superposés : sources ponctuelles sur espace statique (elles ne sont pas dans le même espace)		polyphonie d'espaces
	contrepoint		
	travail sur les attaques		
	espace géométrique		
	événements proches		
	drôle de feeling, le son est tout à gauche (l'auditeur en question est assis à proximité de l'enceinte gauche et ne perçoit pas le côté droit, contrairement aux autres morceaux écoutés)		

Table 2 : collecting (French) words examples: « raw » version and summarized version

The here above corpus of words is part of all the words proposed by teachers and students of the Conservatoire Federal Geneva (May 06), during a Master Class on Space (5 professionals, 12 students in composition).

Words are not translated in English in order to keep all necessary precision.

5.3. The listening sessions

Several listening test sessions took place in front of different kinds of audiences, musically educated, but generally not space experts: instrumentalists, composers, acousticians, sound engineers, students…

- 15th of March 2005: lecture about Space in the composition class of CNSM de Lyon.
- December 2005: "space day" at Music Department / Université Lyon 2.
- 12th of May 2006: Master-class on space at CFM (Federal Conservatory of Music in Geneva), in the composition class. Commented listening of musical works and audio examples. Listening tests "looking for vocabulary allowing to describe space listening". In the presence of sound engineers of Geneva, teachers and pupils from CFM Geneva. Organizer: Emile Ellberger.
- 7th of February 2007: ENM of Villeurbanne, composition and studio technology classes.
- Other private meetings were held in the presence of friends, musicians and composers.

6. EXPLOITING THE LISTENING TESTS RESULTS

6.1. Recopy and cleaning words

Proposed terms were copied just as it is, slightly grouped by similarities. 10 to 20% of responses were suppressed due to off-topic (comments about the work itself or about timbre, poetic description uneasy to exploit…)

As an example, table 2 presents some results. It is a small part of the word collection conducted in Geneva.

6.2. Consolidation of words by "families"

Words presenting similarities are grouped together. These "families" then receive a title (surnames or category title), the best representation of their content.

Some words (or phrases) may appear twice in different families.

Currently, no deletion of words, nor any of rewriting attempt is done (or very few if so…). Some antonym word additions are made: when it clearly appears that a word is cited and that its opposite is not cited.

There is also no attempt to standardize the collected terminology. For the moment, operations simply consists in observations, draft classification and formalization trial.

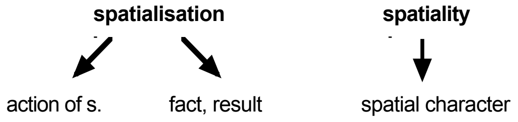
6.3. First analysis of the collected words

This vocabulary consists of collecting **nouns** and **adjectives**. This commonplace distinction will become very important in the following lines.

A) Nouns

The **nouns** describe:

- either a space state or a space situation ; we shall call this space character : "spatiality" (see box below);
- or a "space object", a phenomenon of space, action or the result of an action, which we will call "spatialisation".



It quickly becomes clear that one must distinguish the static situation (a more or less stable state) and the dynamic situation (state change or movement). That means making distinction between:

spatiality or spatial situation	spatiality change or perception situation change
spatialisation or « space object »	displacement of a « space object »

These words refer:

- either to spatial perception of the diffusion place (few occurrences) (see §6.4);
- either to spatial perception of spatilisation sound system (few occurrences) (see §6.4.b);
- either to perception of "spatiality" (most frequent occurrences) (see §6.4);
- or to perception of spatialisation (most frequent occurrences) (see § 6.4.d h).

"Spatiality" seems to be similar to the result of spatialisation perception (spatialisation action), or to the perception of an aesthetic choice (which would be a kind of "intellectual action".
Example: listening to a "soundscape recording" creates a typical spatiality sensation.

b) Adjectives or qualifiers

Adjectives bring several precisions about nouns, so about families. They characterize spatiality, spatialisation, movement, distance, and so on. (see § 6.4.d h).

Adjectives were often cited in particular situations: distant plan, large space, fast ping-pong…

We tried - as a first step – to decontextualize these adjectives (i.e. extract adjectives out of any context), hoping to give them the broadest possible terms. It does not work! A "sound bath" cannot be swift, a "sound plan" cannot be pinpoint or accurate, and so on.

This attempt to generalize made us aware of the necessity of contextualising adjectives and as a consequence of the various nature of nouns describing spatiality. We will come back later on that point.

6.4. Commented presentation of family clusters

As a reminder or as an illustration, some collected words are quoted in the insert and in italics in front of each family.

The arrow ➥ points observations, comments, procedural details on the sidelines of the main speeches.

a) Listening room perception

reverberation, room effect, ambiance
impression of an huge hall
event that sound within that space

Awareness (or not) of the listening room or of the loudspeakers seems rather rare and linked to specific space discourses (such as rapid movements of punctual events, lack of space polyphony, excitation of only one loudspeaker at a time, i.e. a unit space mass space).

b)Sound system perception

We forget loudspeakers
≠ awareness of loudspeakers
The sound travels at the surface of loudspeakers membranes

Once again, awareness (or not) of the loudspeakers existence seems to be linked to the existence (or not) of sound trajectories or movements.

c) Perception of the spatiality of sound events

Through the richness and diversity of this vocabulary spontaneously proposed by hundreds of people in order to describe a dozen of sound examples, five categories seem to appear quite distinctly to describe spatiality: the "sound bath", the "image of space", the "sound plan", the point and the "démixage" (see box below).

categories
sound bath, immersion, ambiophonie, surround, wrapping, holophonic *ambiance, noise everywhere, everywhere,* *we are in the middle of something* *feel like sitting in the middle of an orchestra*
space images, phonographies, landscapes of sound events, sound realism *wide open spaces / closed space ≠ open space* *sound realism ≠ imaginary space*
sound plans, space plans, layers *timbre sound plans, dynamic space plans*
points, pinpoint events, small precise things, pointillist, work by spot, gap between spots
démixage *n rather spot like sources, not mergings*

Table 3 : list of the 5 types of spatiality

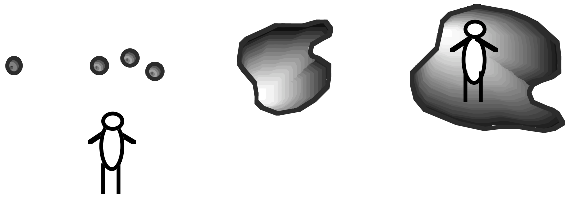


Figure 1 : point, "démixage", plan and "sound bath"

A latter analysis will show that the existence of these five types of spatiality seems to be consistent: the listener perceives one of the following situations:

A. The listener is outside the space area and he perceives:

- one point source: the point;
- n not merging point sources: the "démixage" (see box below);
- a bulky object: the plan (or volume);

B. The listener is within the space area generated by the projection of music and he perceives:

- sound coming from everywhere: the "sound bath";

C .The listener finds space phenomena belonging to life reality: the "space image".

Note: the A and B situations rather seem to be issued from artificial treatments in the studio, whereas the C situation seems to belong to the field of sound realism.

The word « démixage » was defined in the « vocabulaire de l'espace… »[7], as follows:

i) process of not mixing; i.e. setting n "spatially" independent sources at a place (and listening as is) on n channels.

For example, in pop music, instruments are very often recorded one by one and put on separate tracks. Listening to that un-mixed record on n-speaker system is very often interesting (spatially more interesting than the stereo reduction). This listening situation improves intelligibility, comfort, pleasure, but does not generates encompassing or surround space nor space movement, neither space polyphony. The lack of correlation between channels and the "artificial" studio work do not generate a true "real soundscape".

ii) the perceived impression while listening to a multiphonic source whose channels are not spatially correlated and do not fuse.

This process is easily feasible at multitrack instrumental sources recording (or easy to recognize at listening), but it is by no means exclusive of instrumental music.

➥ After the presentation of the whole words families, the following question will appear: is "démixage" a real form of spatiality or is it rather a way of considering polyphony? Some answers we will get later.

Next, let us consider the study of adjectives or qualifiers. As announced in §6.3.b, presenting each family will be followed by a context study.

d) Sound position characterization: localisation

Plans can be: frontal or lateral, forward or backward, left or right, upside or down, in front or behind…
Sometimes it is possible to perceive verticality or elevation.

Contextualisation:

Those adjectives seem to be applicable only to a point or a plan or an "object". They may not apply to a "sound bath", nor to a "soundscape", neither to a "démixage", at least not as a whole. However, they may apply to a pinpoint sound that would be part of a "sound bath", a "soundscape" or a "démixage".

e) Geometry of a "space object" description: shape, size, space encompassing

Points can be spotlike or diffuse, focused or unfocused.
Or dense
One, two or three dimensions

Contextualisation:

Same remarks as above: it only applies to an "object": point, plan…
Note: For further information about the concepts of mass, area, site…, please refer to J.-M. Duchenne's writings [15] ou [7].

f) Distance characterization: close, distant

{point, plan, event}nearby, close, distant
depth, depth of focus,
sound events distributed in space
near and far (at the same time)
events are not in the same space
points in the same plan
points (sound spots) that remain on the same plan

{point, plan, event} getting closer, moving away
with / without depth work
The sound travels at the surface of loudspeakers membranes

wide open spaces / closed ≠ open space

➥ The concept of open or closed space is deliberately separated from the category "listening. place perception". Because listeners do not speak here of the real listening room itself (physical reality), but of the perception of an imaginary listening space (completely independent from the physical location). This notion seems rather close to the perception or the depth of focus notions.

Contextualisation:

Same remarks as above: it only applies to an "object": point, plan…

g) About the movement notion

Many auditors in very different circumstances used the terms: "mobile" or "movement". An attentive second-listening of the concerned excerpts clearly shows the need to distinguish an internal and an external sound mobility.
In the case of **internal mobility**, there is no sensation of movement, or geographical relocation, while **external mobility** is clearly associated with the perception of a movement or with a sound movement (virtual movements, as neither the loudspeakers nor the actual acoustic sources do move).

h) Sound internal movement characterization: "entretien", grain, internal agitation

internal mobility of sound
granular
shimering or sparkling (chatoyant), radiating (irradiant), moving, mobile, bewitching (envoûtant), floating, changing, enveloping
space trill, scintillating
moving everywhere, always in movement

➥ It does not seem that this distinction is linked to trajectory dimensions or spatial cluttering; it seems that these two phenomenons are of different natures. I would venture the following hypothesis : a timbre modification triggering a modification of distance perception or of spatial mass.

i) Movement characterization

fast, slow, accelerating, slowing
discrete, continuous movements, trajectories with or without accidents ≠ fragmented movements
predictible or evident ≠ unpredictible movements
unprobable, disparate, strange, surprising
movements bound to sound materials
dynamic movements, always in movement
stability impression ≠ un-stability

g) Trajectory characterization

points moving inside a plan
points turning, points moving everywhere
music is moving everywhere around us

points into the same plan
points that stay in the same plan
static ≠ dynamic space, fixe ≠ mobile space
mobility
movements, trajectory, space figure
envelopment, encompassing, surrounding
growing, spreading ≠ contracting, squeezing
sound that leaves and reaches a destination
pan, ping-pong, response, joystick, travel
points that turn, rotation, circulary movement of one sound at a time
music is moving all over us
geometric space

Contextualisation :

Once again, these movements can only be applied to points or « objects » less voluminous than the listening room. It has no sense for the "sound bath", nor the "soundscape", neither the "démixage" (except if one only consider individual elements composing them).

j) Space polyphony and depth of focus

At least, the following words characterize spatial superposition or encompassing of several « objects ».

counterpoint, space superposition, stacked spaces, layers
near and far sound (at the same time)
sound bath (with some ponctual sounds)
events in different spaces, sounds ditributed in space
ponctual sources over a static space (not in the same space)
multiphonic discourse
static and narrative events
large variety of effects

« Depth of focus » should be linked to the geometrical occupation of space (in the depth axe): only one sound object doted of an important spatial mass or several distinct sound objects spread over space.

« Space polyphony » underlies something more conceptual, such as simultaneous perception of several spaces or several spatialities or several space discourses.

Both terms partly recover, but are not synonyms.

Contextualisation:

These words do not characterize any of the five space categories, but in fact how several spatialities may combine together.

k) Musical discourse et space discourse suitability

layers crossing space

movements creation by fragmentation
space specific to used sounds
materials influencing movement
figures uncorrelated with sounds, with timbre

➥ It seems that space movements « work better » when coinciding with sound attacks and when sound timbre owns a rich spectrum. Counter-example : a synthesized flute sound generates a strange feeling …

➥ In the same way, hyper-complex and fast movements generate a kind of « stroboscopic » perception of movement.

➥ Reverberation and « distant » effect also jam movement perception.

7. ANALYSIS AND REFLECTIONS

7.1. Relations between nouns and adjectives relations between spatialities and qualifiers

Through the study of our word set, it appears that it would exists:

- 5 spatiality categories,
- 4 or 5 families of adjectives or qualifiers.

they are gathered in the table below.

sound bath	
sound scape	
sound plan	localisation geometry distance internal agitation movement
point	
démixage	

It appears that adjectives or qualifiers only apply to two spatiality categories. This particularity leads us to think that these five spatialities might not be of the same nature.

➥ In a quite similar way, the following sentence poses a problem : « *Les images d'espace peuvent être : narratives, cinématographiques* ». « *Space images can be: narrative, cinematographic* ».These qualifiers are evidently belonging to another level than those displayed above. Probably a metaphorical level

7.2. Different natures of spatiality

The study of this here above table is full of lessons.

a) Finite or infinite encompassing

Spatialities having a finite encompassing would rather receive geometrical-like qualifiers (localisation, geometry, distance, internal agitation, movement…). We shall call these spatialities: « **space objets** ».

Spatialities having an infinite (or huge) encompassing do not seem to own qualifiers. We shall call these spatialities: « **space environments** ».

8. SYNTHESIS TEXT: CARACTERIZING SPACE LISTENING (IN ELECTROACOUSTIC SPATIALISED MUSICS)

As a synthesis of our study, we emit the following propositions:

1) **Space** is the environment in which we are listening to (electroacoustic musics) and in which we can locate « objects ».

2) The **environment** covers the whole space, or at least, such a huge part of it that it becomes difficult to find limits. Environment can emcompass (or surround) the listener (**sound bath**) or be external of him (**plan**, **soundscape**) ;
Soundscape or sound image notions refers to sound reality or sound realism.
This perceptive environment is generated by a sound system located into a listening room.
The perception of this spatial environment (or **spatiality**) can be independent (or not) of the listening room; The sound system can be perceptively « transparent » or « revealed ».
being aware (or not) of the listening room;
being aware (or not) of the recording room;
being aware (or not) of the sound system.

3) « **space objects** » occupy a finite portion of space.
The space part occupied by a "space object" is called the **area** (l'**étendue**). Several **dimensions** allow to measure this area: volume, size, length, wide, height, depth, *etc.*
The area of an object can be idealized: point, line, surface, volume.
Localisation is the fact of situating an « object » in a **place** or its relations to a specific environment or another object..
Orientation allows to situate objects in relation with other objects, according to special relatiopnships axes **verticality**, **horizontality**, **frontality**, **laterality**.
frontal or lateral, forward or backward, left or right, upside or down, in front or behind…
spotlike ≠ diffuse, focused ≠ unfocused, dense

4) **Distance** is the interval that separates two « objects ».
near, far

5) If an « object » is time dependant, it becomes a **space event**.
Movement is a space position change of an « object ». This event lasts a certain amount of time. There are several types of movements:
(a) the **internal movement** of an object;
(b) the **deformation** of an object;
(c) the **displacement** or location change.
The movement nature can be made more explicit thanks to several characteristics:
fast, slow, accelerating, slowing movements,
discrete, continuous movements, trajectories with or without accidents ≠ fragmented movements
predictible or evident movements ≠ unpredictible movements
bound to sound materials
dynamic movements
always in movement
If coherent and predictable, a movement can be idealized:
Some movements or displacements can be identified and formalized.
line, pan, rotation, zig-zag, expansion, contraction…

6) **Space polyphony** or **space superposition** :« Objects », « events » and environment(s) can combine together, without merging (i.e. staying spatially distinct one from each other).
Depth of focus allows to describe superposition of several objects or events —

or of one object (or event) in relation to its environment — in the depth direction.
Cohabitation of several motionless « objects », set all around the listener and not merging together, is called **démixage** (referring to some studio practices). A more explicit word would better be found in the future.

9. CONCLUSIONS

The aim of the present study is to clarify vocabulary of space perception in electroacoustic musics (composed ou spatialised in pentaphony): a link between words and sounds. To do so, a set of listening tests, processes and word collection have been developed and realized between 2004 and 2007 on hundreds of people. These first tests and word collection appear to be really interesting and fruitful: a great amount of crosschecking information has been gathered.

A first classification was realized, separating nouns from adjectives, proposing five types of spatialities and about half a dozen of qualifier families.

An analysis of the relations between nouns and adjectives, as well as a study of the adjectives contextualisation allowed us to clarify the situation and to propose a synthesis text of our whole observations.

In order to consolidate those first observations and to refine this vocabulary, other test sessions should be realized; probably with new sound examples specially realized on purpose.

Perception description of space in spatialised electroacoustic musics now owns an embryonic lexicon, written words that attempt to describe the listening spatiality. This « writing process » — probably unperfect or uncomplete — is however fundamental for further communication and reflection.

For over 60 years, composers have been putting electroacoustic music in space. However, very few documents describing space composition techniques (by instance: [1], [2], [10]) or spatialisation methods or spatialisation gestures [13] have been elaborated and published. But it seems that nobody ever tried to really describe and formalize spatial listening processes. That is now done!

As already said in the introduction, fixing spatialized musics on a consumer multichannel support (CD or DVD) and proposing a spatial listening vocabulary might be at the origin of a new discipline: space musicology!

10. REFERENCES

10.1. Books and papers

[1] DUCHENNE Jean-Marc, *Des outils pour composer l'espace*, Actes des JIM 05, MSH / Université Paris VIII, mai 05.

[2] DUCHENNE Jean-Marc, Pour un art des sons vraiment fixés, *in Ars Sonora*, No. 7. Paris: Ars Sonora/CDMC : 36-68, 1998. (URL : http://www.ars-sonora.org/)

[3] HAIDANT Lionel, *Prise de son et mixage en surround 5.1*, éd. Dunod, 2002.

[4] LEROT Jacques, *Précis de linguistique générale*, Les éditions de minuit, 1993.

[5] MENEZES Flo, « La spatialité dans la musique électroacoustique. aspects historiques et proposition actuelle », *L'espace : Musique / Philosophie*, Textes réunis et présentés par Jean-Marc Chouvel et Makis Salomos, Coll. Musique et Musicologie, L'Harmattan, 1998

[6] MERLIER Bertrand, *Vocabulary of space in electroacoustic musics: presentation, problems and taxonomy of space*, Actes du colloque EMS (Electronic Music Studies), Pékin (Chine), oct 06.

[7] MERLIER Bertrand, *Vocabulaire de l'espace en musiques électroacoustiques*, coll. Musique et sciences, éditions Delatour, France, 2006.

[8] MERLIER Bertrand, *Vocabulaire de l'espace et de la spatialisation des musiques électroacoustiques : Présentation, problématique et taxinomie de l'espace*, Actes des SMC 06 (Sound and Music Computing) / GMEM Marseille, mai 2006.

[9] MERLIER Bertrand, *Réflexions à propos de la mise en espace de la musique électroacoustique dans les logiciels audionumériques*, Actes des JIM 05, MSH / Université Paris VIII, mai 05.

[10] MERLIER Bertrand, *Surround, Mode d'emploi*, revue « les cahiers de l'ACME », n° 221, fév. 2005.

[11] MERLIER Bertrand, Musiques électroacoustiques mises en espace pour le *surround* 5.1 et encodées en dts. *Actes du colloque JIM 2000* (Journées d'Informatique Musicale), le 18 mai 2000 à Bordeaux au SCRIME - ENSERB.

[12] MERLIER Bertrand, À la conquête de l'espace, in *Actes des Journées d'Informatique Musicale.* p. D1-1 à 9, publications du CNRS-LMA, n°148, MARSEILLE, ISBN : 1159-0947 1998.

[13] VAN DE GORNE Annette, L'interprétation spatiale. Essai de formalisation méthodologique, revue DEMéter, Université de Lille-3, déc 2002 (disponible en ligne : http://demeter.revue.univ-lille3.fr/interpretation/vandegorne.pdf).

10.2. Web sites

[14] GETEME, http://geteme.free.fr

[15] Duchenne Jean-Marc : http://multiphonie.free.fr

[16] Merlier Bertrand, http://tc2.free.fr/Espace/

[17] Thélème Contemporain, http://tc2.free.fr

[18] Swedish Radio – Multichannel Sound 5.1, http://www.sr.se/cgi-bin/mall/index.asp?ProgramID=2446

[19] TELARC, http://www.telarc.com/dts/ propose une petite collection de CD en DTS 5.1 ou DVD de tout genre : chansons, pop-rock, classique…

[20] http://www.cddts.net/

[21] http://cddts.free.fr/ offre un petit tutorial qui vous permettra de réaliser assez facilement des CD Audio 5.1 DTS à partir de vos MP3 favoris, un chat, des liens, ainsi qu'un annuaire de ceux qui pratiquent le DTS.

[22] http://www.5dot1.com/

[23] http://www.5dot1.com/equipment/ac-3_and_dts_software_encoders.html

[24] Site officiel du DTS, http://www.dts.com

10.3. Discography

[25] *Musiques électroacoustiques spatialisées en 5.1 et encodées en D.T.S. (vol. 2)* (Barrière, Bayle, Bouttier, Diennet, Duchenne, Favre, Merlier, Risset)*,* Thélème Contemporain , CD 14, 2004. http://tc2.free.fr/CD51.html

[26] *Musiques électroacoustiques spatialisées en 5.1 et encodées en D.T.S. (vol. 1)* (Merlier Bertrand : « Picson, le hérisson », « Nébuleuse NGC 2359 », « Les chevaux de Ladoga », « Sillage »)*,* Thélème Contemporain , CD 11, 2000. http://tc2.free.fr/CD51.html

[27] ORTI Guillaume & SENS Olivier, *Reverse*, Ed. Quoi de neuf Docteur, 2005 (www.quoideneufdocteur.fr)

[28] EAGLES, *Hotel California*, in DVD promotionnel #4 édité par DTS USA, 1999

[29] Swedish Radio – Multichannel Sound 5.1, http://www.sr.se/cgi-bin/mall/index.asp?ProgramID=2446

Sound as *Multiplicity*: Spaces and representations in computer music composition

Arturo Fuentes
Paris VIII University/CICM (Centre de recherche Informatique et Creation Musicale), Paris, France
af@arturofuentes.com

Abstract — **This text deals with the subject of sonic spaces within the field of computer music composition. Highlighted by the notion of multiplicity, the sound will be analysed as a *multi-representational space*. This central idea will take us to consider some proposals of the hermeneutical criticism of representation, where we'll observe the emergence of sonic spaces from an *action-perception* perspective: our musical significations appear at the very moment we execute a "local action" in the composition process. Multiplicity is produced by singularities as well as singularity is conceived as a multiple entity: depending on our *operatory procedure in music composition* we shall consider a sound as One or as Multiple. In music composition, human-computer interaction moves towards this problematic.**

I. Introduction

It is important to make clear that in this article the notion of multiplicity [1][2] is not employed as an adjective in the way it is used in speech, for example, when one says: a multiplicity of musical instruments, a multiplicity of styles, etc. Instead we are focusing on a "strong sense of this word" [3], as Deleuze has stated. The French philosopher refers directly to the *"Données Immédiates"* [4] of Bergson, where we find that "a number is a multiplicity", which is not the same thing as a "multiplicity of numbers". This article aims to explore the next idea: "a sound is a multiplicity", it also proposes that sonic interactions in a musical work create a "multiplicity of spatialities". We are thus considering a compositional strategy in computer music which recognizes the concept of sonic space not only as the physical medium where the sound is diffused, but also, as an 'operatory' category (*catégorie opératoire*) which outlines many other "composable spaces" (metaphorical or representational) every time that we interact with the computer.

II. Composable Space

In the field of computer music composition we can see that the concept of "sonorous object" (*objet sonore*), reelaborated by the composer and theoretician Horacio Vaggione in many of his writings, keeps the same "strong sense" of multiplicity, as proposed by Deleuze. Vaggione develops the concept of musical space as a "composable space" [5] based on the emergence of musical meanings revealed in every single act of composition. Following the same line of thought, Vaggione assumes that the composer must deal with a multi-representational space contained in the alphanumeric environment of the computer, where an operatory category, that of the sonorous object is defined as a "multiple". More precisely, the sonorous object constitutes a "multiple unity" [*ibidem*].

The composable space is considered as an element that can be articulated; it is then conceived as a material, Vaggione affirms: "If the space is conceived as compositional material, that means that it is essentially a space of relations" [*ibidem*]. We could also visualize the computer itself as a "multiple unity": we understand that it is our interactive practice which generates a multiple sonic space. However, we could not musically evaluate the computer in terms of its calculation processes or in terms of its musical representations, because, as we know, a "musical space" does not exist in it; we *compose* this space of relations following our aesthetic assumptions, actually, we need to "musicalize the computer" [6], as proposed by Risset.

It should be clear that the composable space is made of a conjunction of operations producing musical significations. In computer music, these operations (producing musical meanings) can be found being part of both spaces, the real and the virtual; Risset has established that "the notion of space is consubstantial to the electroacoustic music […] it provides us a real physic [space] to play with the virtual" [7]. As we can see, the sonic space results from the interaction between the real and the virtual space, it is a composed element, a "multiple unity", as we have seen before. We could not certainly evaluate musically the diffusion of a sound in a physical space without a certain amount of "compositional background" (metaphorical, structural, etc.). It would be also difficult to musically think of a space represented in the computer without attesting it in a physical space. The sonic space is made of a connection between the real and the virtual space; according to Risset: "When we create a virtual space, we create a simulation of propagation in an acoustic space" [*ibidem*].

III. Spatialities

In connection with the arguments of the hermeneutical criticism of representation, which will be developed later in this text, Vaggione affirms that an object is part of a network of objects, while at the same time containing its own network [8]. It is from this perspective that we can think of a sound as a multiplicity. This concept of the sonorous object, which is defined by the research field that concerns itself with the *operational procedures in music composition*, comes closer to the concept of sonic space as developed in this article.

We thus consider the musical work as a sonic space where a "multiplicity of spatialities" emerges. In order to clarify this idea, we can refer to the sonorous object defining its own space of operations, but being, at the same time, part of a group of operations contained in the space of a musical work. In music composition a "multiple unity" is a sonorous object, but also, it could be a piece of music, so to speak a morphology, which in every case contains multiple *operations*. The notion of *sonic space* is thus considered as an "operatory category" located in the micro and the macro level of sound, it is something that we use to compose our "musical distinctions", pointing out the difference between the layers of time in our music: "The musical space is something to be composed" [cf. 5], as has been clearly stated by Vaggione.

What we're looking for as composers when working with computers is a "composable space", more than a "represented space" (an *a priori* representation of a sonic space where, hypothetically, we would develop a thinking about the objects contained in it). In a composable space the objects themselves put forward a spatial thinking, we can see that their forms result from their operations; in fact, as it has been proposed by Granger: "object's space and operation's space are reciprocal" [9][10]. We employ an interactive approach when we claim this reciprocity between the objects and their operations. Under this perspective, the space of the musical work is a particular kind of object giving us the possibility to compose the relations of objects during a temporal flow. We shall recognize a sonic space as defined by an ensemble of operations situated at different layers of time.

IV.Composable Distinctions

In computer music composition, we are focusing on an interaction of a multiplicity of musical spaces of many kinds (physical, operational, perceptual and metaphorical). We thus understand the composition of the musical work as a "multiplicity of composable spaces". To paraphrase Deleuze, it could be said that we are distinguishing "the kinds of multiplicity" [cf. 3]. Certainly, this statement sets us apart from a dialectical position between the One and the Multiple, and the composer is thus engaged in differentiating the levels between them: one musical figure that could be perceived as unitary in one representational scale, could be defined as a multiple in another. As Vaggione has put it, as composers, we are led to make "composable distinctions" [cf. 5]. The numerical field has allowed the composer to work in this sense, as we know, the alphanumerical code has opened more possibilities for the composition of the sonic material, revealing to us the finest differences contained in sound.

The musical syntax has been increased because of this micro-temporal manipulation of sound; we thus distinguish several concepts close to this multiplicity, as we have seen before, that of the sonorous object could be applied to all temporal sizes, marking out a multiple space in a musical work. Even if we are not establishing a dialectical position between the One and the Multiple, it is important to observe that in order to articulate a multiple space which conveys all composable distinctions, we need to set (encapsulate) the connections between the operations. Vaggione makes the following statement concerning this aspect of the multiplicity of the sonorous object: "In the informatic jargon, the *encapsulation* term corresponds to a linkage of an ensemble of proprieties and behaviours pointing out the creation of an object" [11]. The encapsulation procedure furnishes a singular attribute to the work, some kind of artistic quality which marks out a musical context. We can think in a composable space as a multiplicity of encapsulated entities, which is underscored by Wittgenstein's remark establishing that "the configuration of objects produces state of affaires" [12]. It is important to mention that a sonic space conceived as a state of affaires, would mean that we are engaged with a permanent critical point of view vis-à-vis the sonic relations, definitely, this alert attitude allows the composer to establish the different degrees of its sonic material, which could go from the One to the Multiple.

V.Proprieties and Behaviours

As it has been analyzed in a precedent article [13], it should be clarified that a computer cannot create music by itself; however, it can help us generate musical ideas when we introduce a *behavioural specification* into it: by interacting with users, other computers or physical systems. One special feature of multiplicity is underlined when we approach a behavioural specification: the *reduction* of contents when the composer works (*designs*) with the representational environment of the computer.

Music creation with the computer conveys an irreversible process of time; it changes at every step of transformation. The composer cannot determine all the representational features of sound that would correspond to his aesthetical assumptions, the sonic space of the piece is something that emerges by an action-perception perspective thanks to a selection of sonic elements that constitute local musical significations. In the line of Granger's ideas we would say that the composer locates "knots of new significations" [cf. 9, p. 389] in the sonic space of the piece. Briefly, the composer creates a musical representation in the computer that makes *sense* to his ears.

There is a connection between the *action* of creation in a representative space and the *perception* of these acts. We deal with the operations of two kinds of sonic spaces in order to create a musical signification. In composing with computers, we have a tendency to randomly multiply the sound transformations in our imagination (our musical ideas are normally formless), however, it would be better to consider this cognitive uncertainty as an inevitable aspect of music invention. Thus we use the computer to *unify* this *multiplicity* in a composition strategy. In keeping with the words of Winograd (in the informatic field) who stated that "Design is by nature both holistic and ruthlessly simplifying" [14], we shall then introduce one of the central lines of thought of this text which is that the *design* of sonic spaces is both holistic and simplifying. We use the computer to *unify* or to *multiply* a sonic space.

It could be stated, paraphrasing Risset, that when we design a sonic space we create *simulations* of propagation of sound as well as *simulations* of the representations of this propagation, in fact, we have a reduction of contents in both simulations. Those reductions of contents are directly influenced by our musical practices which deal with our aesthetic postulations. Following the words of Winograd we clarify the remarks presented here: "A designed artifact, whether it is a piece of communications software or a city park, must address the complex mixture of human needs, embodied in a weave of physical and social interaction. But the design itself cannot embody all of these complexities if it is to be constructible and understandable. The design must embody a simplification, leaving room for the texture of the world to be filled in by the interpretation and practices of those who use it" [*ibidem*].

The design of a sonic space must also embody a simplification of its multiplicity in order to be constructible and understandable. In electroacoustic music, when several variables are engaged in the process of composition (signal treatment, spatialization process in real time, multi-channel recorded 'tape', musicians playing), it is better to follow this suggestion. In complex systems of music production, when this simplification is not apprehended, it could result, for example, that spatial treatments of sound will not be perceived: too many movements of sound between the speakers will form a confused texture which would cancel a detail of the spatialization, as it could be, for example, the directional trajectory of sound covering the physical space.

Certainly, it should not be understood that multiplicity has "simplest" audible results only. What we have been talking about refers directly to a creation of a composition strategy that should be explicit in its own terms and operations. It conveys a design following the proprieties and behaviours of a sonic space involving the composer and the machine. Multiplicity in a sonic space is revealed by its singularities, however, a compositional problem appears: how can we compose a multiplicity without loosing a detailed perception of singularities?

VI.Singularities

We have been referring to simplification within multiplicity. Thus it is interesting to reflect upon how much our musical strategies are transformed when we direct towards complexity, which at a first sight would have more points in common with the concept of multiplicity. Singularity, nevertheless, unfolds other questions about the construction of sonic spaces.

In an interactive situation involving the notion of multiplicity, the computer can be considered as an *integration interface* where several compositional operations are symbolically formulated in non-dialectic terms (composer-computer). As composers we situate the representational space of the computer in an artistic perspective: it is a compositional tool and a musical instrument. We perceive it as a *spatial interface* which gives us the possibility to codify musical (compositional) ideas within a representational environment. We also view it as highlighting the dynamic nature of our musical practices: "it offers us the possibility to transfer *abstract-artistic* to *symbolic-algorithmic* information" [15].

The computer can help us to multiply the transformational possibilities of sound when we "discover" its representational space, which is made up of many symbolic layers of sound. This space, or network, could thus be seen as a *multiplicity* which is constructed out of an infinity of interconnected *single acts* (these acts are referred to as *singularities*, due to the *aspectual* and *qualitative* features of sound that interest the composer during the process of its transformation in a computational environment).

The present text could also have been untitled *"Sound as Singularities"*. It is clear that the sound holds a multiplicity made of an ensemble of singularities. Thus it is important to consider that the representational space of the computer allows the composer to go deeply into the sound, searching for details. The composer can now articulate these micro-elements and change of temporal scale at every time of the process: "The computer is an ideal tool that allows us to deal with this situation", affirms Vaggione, and he continues: "With this tool we can reach any level of operation and explore all the desired and possible links between different levels. It is true that we are forced to use different systems of representation, choosing the ones more adequate to each particular level. This is why we are confronted with disjunctions and nonlinearities; a symbolic system that describes well a given morphology at a particular level can become nonpertinent when applied in another level" [16]. Those disjunctions and nonlinearities create singularities in the sonic space, as composers we are confronted to particular cases at every level of representation. Multiplicity is formed by the ensemble of particular cases in our process.

VII.Synoptic View

Multiplicity reveals the diversity of elements contained in a unity, outlining a contextual vision of the ensemble. We thus attain a "spatialized vision" of details contained in a sound. In these circumstances, we do not lose our approach to the qualities of sound, they are, so to speak, projected in a multiple space, amplified or disturbed by the contact of the ensemble. Singular elements (individualities) presented in the multiplicity attain other sonorous qualities in their macro-morphological arrangement. According to Morin: "The organization of the whole is something more than the addition of the parts, because it uncovers the qualities that would not exist without this organization" [17]. As composers we are thus interested in the "emergent qualities of the singularities".

We compose a dimensional space within multiplicity, our perception goes around the ensemble of singularities. This exploration of the multiplicity is confronted to heterogeneity, however, our listening, as well as our view, has tendency to rebuild regularity in heterogeneity: "Things that resemble each other are tied together in vision", has stated Arnheim [18]. In music composition, this proposal takes us to reflect on the qualitative emergence of the sonic space when multiplicity is constructed by the connection between sounds and not by connecting "term to term", which would be a restricted linear-logical construction of our musical structures. As composers we strive, following Wittgenstein's remarks, "not after *exactness,* but after a synoptic view" [19].

VIII.Critical Remarks

In order to clarify the action-perception remarks exposed in this article, an approach to the hermeneutical criticism of representation is exposed here. We need to consider the texts which have been published since the end of the 1970s up to the present time in the field of informatics and cognitive science (offering some alternative propositions to classic cognitivism). Of particular interest for this critical position are the proposals of Brooks [20][21]; Chapman [22]; Dreyfus [23][24], Wegner [25]; Winograd [26]; Winograd and Flores [27]. The argument developed by these authors has been clearly formulated by Vaggione: "representations do not have an intrinsic reality, they are tools pointing out a contextual emergence, corresponding to a *situation*" [cf. 8]. We, of course, understand that sonic spaces do not have an intrinsic reality either. The notion of *interaction* is then centred in the midst of the composition process, avoiding the predictability of closed systems and reductive formalisms.

Vaggione employs an interactive approach when he claims that "Operatory representations in music could not be identified as representations of "mental processes": Even if the musician keeps and deals with quantities of musical relations (defined as elements out-of-time) in his mind, these elements will always be part of an "external world", that of the music" [cf. 8]. This argument in the field of music composition evokes the work of Winograd in the field of computer theory in which several theories regarding a hermeneutical approach to informatics have been developed. He recognizes that "the concepts emerge in the interaction more than in the machine or in the head of the user" [28].

This critical position that we have classified as an operatory procedure in music composition maintains that musical ideas result from human interaction with the computer, similar to how one interacts with a musical instrument or with a sheet of paper on which one organises the notes. As a criterion for compositional position, musical ideas are thus intended as sonic operations represented in a configuration system (a *composable space* as Vaggione has stated). It is clear that musical ideas constitute one more space which should not be considered in dialectical opposition to the representational space of the computer.

Physical spaces as well as representational spaces produce compositional operations where music emerges. In the line of Di Scipio's ideas, Solomos affirms: "The concrete space –the place– is part of the music to be composed. This aspect makes the music itself an emergence" [29]. This last sentence confirms one of the central ideas of this article: sonic spaces (physical or representational) are operatory categories in music. According to Nono "the sound reads the space" [30], this would also be understood as an operatory procedure because we have to deal with the qualities of a space in order to produce music: "the space could be "morphophoric", it could serve to specific musical forms" [cf. 7], has stated Risset.

We compose the space at the same time as the space recomposes musical structures, we turn around this interactivity. We are part of the composable space, we are an interactive singularity of the space, integrating all the elements suspended in it. It could be then said, in the sense of Merleau-Ponty, that we are there "as a point or level zero of the spatiality" [31], producing musical operations.

References

[1] A. Fuentes, *Five proposals for contemporary music composition (lightness, quickness, exactitude, visibility, multiplicity).* Doctoral thesis, Paris VIII University, Paris, 2008.

[2] A. Fuentes, *Multiplicité homme-machine: composer avec l'ordinateur,* IRCAM, Séminaire MaMuX, Mathématiques, musique et relations avec d'autres disciplines, published on line [http://recherche.ircam.fr/equipes/repmus/mamux/].

[3] G. Deleuze, "Théorie des multiplicités chez Bergson," conference given at the Paris VIII University, Vincennes, published on line [www.webdeleuze.com], Paris, 1970.

[4] H. Bergson, *Time and Free Will: An Essay on the Immediate Data of Consciousness*, Dover Publications, 2001.

[5] H. Vaggione, "L'espace composable. Sur quelques catégories opératoires dans la musique électroacoustique," in *L'espace: Musique/Philosophie*. Ed. L'Harmattan, Paris, 1998.

[6] J-C. Risset, "Composer le son: expériences avec l'ordinateur 1964-1989", in *Contrechamps* No. 11, *Musiques Électroniques,* Ed. L'Âge d'Homme, Genève, 1990.

[7] J-C. Risset, "L'espace et l'électroacoustique", in *L'espace: Musique/Philosophie,* L'Harmattan, Paris, 1998.

[8] H. Vaggione, "Composition musicale: représentations, granularités, emergences," in *Intellectica,* CNRS, Centre National du Livre et la Maison des Sciences de l'Homme Paris-Nord, Paris, 2008.

[9] G-G. Granger, *Formes, Opérations, Objets,* Vrin, Paris, 1994.

[10] G-G. Granger, *La pensée de l'espace,* Odile Jacob, Paris, 1999.

[11] H. Vaggione, "Objets, représentations, opérations", in Revue Ars-Sonora, No. 2, 1995. French version of "On object-based composition", Composition Theory, Interface, Journal of new music research 20 (3-4), pp. 209-216, O. Laske (Ed.), 1991.

[12] L. Wittgenstein, *Tractatus* [1918], Gallimard, Paris, 1961.

[13] A. Fuentes, "Sonic/Musical ideas: compositional remarks on computer music", in Proceedings of the *Computers in Music Modeling and Retrieval Conferences* [CMMR], Re:New, Digital Arts Forum, Copenhagen, 2008.

[14] T. Winograd, "Designing a new foundation for design", in *Communications of the ACM,* vol. 49, No. 5, Stanford, 2006.

[15] A. Fuentes, "Den Klang hörend komponieren: Der Computer als kompositorisches Dispositiv", in *Dispositiv(e),* Positionen-Beiträge zur neuen Musik (74), Berlin, 2008.

[16] H. Vaggione, "Composing with Objects, Networks, and Time Scales: An Interview with Horacio Vaggione", in *Computer Music Journal*, 24:3, pp. 9-22, MIT, 2000.

[17] E. Morin, "Le défi de la complexité", in *Revue Chimères* No 5/6, Paris, 1998.

[18] R. Arnheim, *Visual Thinking,* University of California Press, London, 1969.

[19] L. Wittgenstein, *Zettel* [1945-1948], University of California Press, London, 1967.

[20] R. Brooks, "Intelligence without representation", *AI Rapport*, MIT, Cambridge, Massachusetts, 1987.

[21] R. Brooks, "Intelligence without reasoning", *AI Memo* 1293, MIT, Cambridge, Massachusetts, 1991.

[22] D. Chapman, "Vision, Instruction and Action", MIT, Cambridge, Massachusetts, 1991.

[23] H. Dreyfus, "What computers can't do: A critic of Artificial Reason", Harper and Row, New York, 1979.

[24] H. Dreyfus, "From micro-worlds to knowledge representation: AI at an impasse", in *J. Haugeland* (ed.), *Mind Design*, MIT, Cambridge, Massachusetts, 1981.

[25] P. Wegner, "Why interaction is more powerful than algorithms", in *Communications of the ACM,* 40(5), New York, 1997.

[26] T. Winograd, "The design of interaction", in P. Denning and B. Metcalfe (eds.), *Beyond Calculation, The next 50 years of computing*, Springer-Verlag, Berlin, 1997.

[27] T. Winograd and F. Flores, "Understanding computers and cognition", Ablex press, New Jersey, 1986.

[28] T. Winograd, "Heidegger and the Design of Computer Systems", in *Conference on Applied Heidegger*, Berkeley, 1989.

[29] M. Solomos, "Notes sur la notion d'"émergence" et sur Agostino Di Scipio", *Actes des Journées d'Informatique Musicale,* Paris, 2005.

[30] J. Dautray, "Une hétérotopie musicale: la collaboration entre Renzo Piano et Luigi Nono sur Prometeo", *Revue Rue Descartes* No. 56, Collège International de Philosophie, PUF, Paris, 2007.

[31] M. Merleau-Ponty, *Sens et non-sens,* Gallimard, Paris, 1996.

Sound Spatialisation, Free Improvisation and Ambiguity

James Mooney*, Paul Bell†, Adam Parkinson†
*Culture Lab, Newcastle University, Newcastle upon Tyne, UK
†International Centre for Music Studies, Newcastle University, Newcastle upon Tyne, UK

Abstract—**This paper documents emergent practice led research that brings together live sound spatialisation and free improvisation with digital tools in a performance context. An experimental performance is described in which two musicians – a turntablist and a laptop performer – improvised, with the results being spatialised via multiple loudspeakers by a third performer using the Resound spatialisation system. This paper focuses on the spatial element of the performance and its implications, its technical realisation and some aesthetic observations centring on the notion of 'ambiguity' in free improvisation. An analysis raises numerous research questions, which feed into a discussion of subsequent, current and future work.**

I. Introduction

The aim of this research is to explore the use of the multi-loudspeaker sound spatialisation system as an instrument in free improvisation. As a starting point, the authors staged an experimental performance, a video of which is available online [1]. The performance raised issues relating to ambiguity in musical performance, as well as technical and aesthetic issues pertaining to the practice of sound spatialisation itself. The purpose of this paper is to document the performance, summarise observations and pose research questions, preparing the ground for further research.

II. Background

A. Sound Spatialisation

In electroacoustic music, sound spatialisation - often referred to as 'sound diffusion' in this context - describes the act of presenting music from CD, audio file, or other fixed medium, to an audience via multiple loudspeakers. The performer controls the distribution of sound among the loudspeakers by way of a diffusion system, often based around an audio mixing desk. As a simple example stereo sound from CD might be spatialised via four pairs of loudspeakers, with one mixing desk fader controlling the level of each loudspeaker. Bespoke systems have been developed by many institutions specialising in electroacoustic music (see refs. [2] through [11]).

In general, sound diffusion practice is applied in the performance of predetermined music, either from fixed medium and/or from a score; in either case the sequence of sonic events is essentially known in advance. It is more rare for this kind of practice to be applied in the context of free improvisation.

B. Free Improvisation and Ambiguity

The outcome of a free improvisation is not known prior to any given performance. The expression 'ambiguity,' broadly, refers to the unknown, and to the experience of 'not knowing,' in a musical context. Similarly, when one hears a sound, one might know (be able to identify) its source, or one might not know, in which case there is an ambiguity. More broadly, in the context of a musical performance one might know what is going to happen next, or one might not know. Furthermore, what is ambiguous for an audience member may not be so for a performer, or *vice versa*. Clearly ambiguity is at work on many different levels within freely-improvised performance, and fostering it can become a creative strategy and catalyst for extended musical dialogues. For Gaver, Beaver and Benford ambiguity 'is a resource for design that can be used to encourage close personal relationships to systems' [12]. Unintentional sounds ('Where did that come from?') can also elicit unforeseen responses as performers evaluate, interpret, and feedback into the situation.

Through the utilisation of electronic technologies the legibility of gesture is often obscured: there may be no direct correlation between input gesture and output sound; the relationship between the two is ambiguous. However, connections may appear legible through exaggerated theatrics. This is particularly evident in DJ practice, where allowing pre-recorded sounds to play unmediated as opposed to physically intervening in them is inherent, and selection and performance are coterminous. A comparable scenario exists in sound diffusion, where 'fixed' compositions are played without any physical sound-generating process on the part of the performer. Used in free improvisation, the selective use of pre-recorded materials can be employed as a means to probe, provoke and generate creative response, at the same time problematising the conventional wisdom that 'music-making skill paradigmatically requires the immediate causal intervention of the player' [13]. Ambiguity is discussed further in [14]. As we shall see, use of the sound spatialisation system as an instrument in a free improvisation engages with these issues of ambiguity on various levels.

III. Dead Dialogues: An Experimental Performance

An experimental performance entitled *Dead Dialogues* was staged at Culture Lab, Newcastle University on 10th March 2008 [15]. Two improvisers – a turntablist and

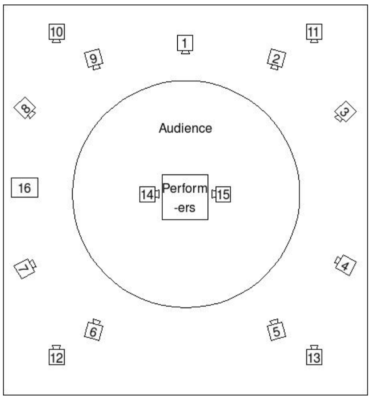


Fig. 1. Venue schematic of *Dead Dialogues* performance.

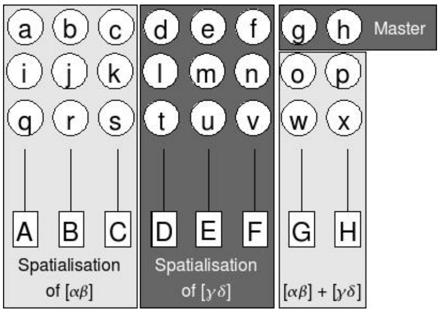
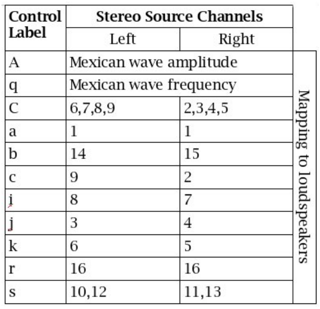


Fig. 2. Diagram of the MIDI controller interface used to control spatialisation.



Control Label	Stereo Source Channels		
	Left	Right	
A	Mexican wave amplitude		Mapping to loudspeakers
q	Mexican wave frequency		
C	6,7,8,9	2,3,4,5	
a	1	1	
b	14	15	
c	9	2	
i	8	7	
j	3	4	
k	6	5	
r	16	16	
s	10,12	11,13	

Fig. 3. Partial scheme for the mapping of interface controls and source-to-loudspeaker routings. 'Mexican Wave' refers to a semi-automated spatialisation behaviour; see [**?**].

a laptop performer – play with a third performer – the *spatialist* – operating the spatialisation system. The sounds generated by the turntablist and laptop performer are spatialised independently in real time by the sound spatialist; however, the spatialist cannot directly generate sound. Conversely, the turntablist and laptop performer are able to select their own sonic materials but have no direct control over the spatialisation. The three performers were located at the centre of the performance space with audience members, facing inwards, surrounding them on all sides, as illustrated in Figure 1.

A. Loudspeaker Array

An array of sixteen loudspeakers was deployed. This comprised nine Genelec 8050A loudspeakers (numbered 1 to 9 in Figure 1), two smaller Genelec 8040A loudspeakers (numbered 14 and 15) positioned on either side of the central performance area pointing inwards, four EAW NT26 PA cabinets (10 to 13) suspended from rigging around 3 metres above floor level, and a single subwoofer (16). The suspended PA cabinets were angled to point straight forwards rather than downwards towards the audience, providing a greater sense of height. Photographs of the setup are available online [16].

B. Spatialisation System

Sound was spatialised using Resound, a real-time, multi-channel, multi-loudspeaker spatialisation system based on freeware open-source software. Briefly, Resound allows the user to control an audio mix matrix using a MIDI or OSC control interface. Matrix nodes can be controlled individually or in groups, with multiple assignments being summed additively or subtractively by the Resound software. In this way, any input channel can be mixed to any loudspeaker in real time, the mapping of matrix nodes to controls having been defined in advance by the user. The system itself is described more fully elsewhere [3][17][18].

The spatialisation was controlled using a Waveidea Bitstream 3x MIDI controller. Eight faders plus twenty-four of the rotary controls were used during the improvised performance. Figure 2 shows the MIDI controller interface schematically, while Figure 3 exemplifies how the interface was configured to spatialise the two stereo sound sources independently among various loudspeaker sets within the array. For example, rotary control *c* (referring to Figure 2) controls the level of a stereo source spatialised to loudspeakers 9 and 2 (referring to Figure 1). Control *s* would spatialise the left channel to loudspeakers 10 and 12 and the right channel to 11 and 13. Referring to Figure 2 it can be seen that the first bank of three faders and rotary controls was used to control the spatialisation of the live electronics as explained in detail in Figure 3. The next bank of three provided exactly the same functionality for the stereo feed generated by the turntablist. The final bank of controls treated the total four source channels as a single source.

IV. Research Questions

Following the performance, and against the background described previously, two emergent research strands are apparent, one focusing on musical issues from an aesthetic, experiential, interpretative or philosophical perspective, the other concerning technique, human-computer interaction and design. Two broad research questions are as follows.

- What can be learned from the use of sound spatialisation as an active instrument in free improvisation? How does the delegation of spatialisation to a third performer impact on the way the performance is experienced by players and audience? What role does ambiguity have to play?
- What HCI demands does the free improvisation scenario place upon the spatialisation system? How can

these issues be addressed through software and hardware design?

V. Some Preliminary Observations

A. From the Spatialist's Perspective

As both the laptop performer and the turntablist use prerecorded materials, with the laptop performer often sampling and processing the turntablist, the true origin of each sonic event becomes unclear. From the spatialist's perspective, it was sometimes difficult to differentiate between sounds originating from the turntables, and those resulting from the live electronics processing (once again, 'Where did that come from?'). The spatialist and the audience can only rely on the gestures of the other musicians to establish (and perhaps misconstrue) sound sources in the piece. Legibility of gesture, and its relationship to human-computer interaction, will be discussed again later.

Furthermore, the dexterity and concentration required to simultaneously spatialise two independent sources was found to be challenging, indicating considerable scope for virtuosity with practice. This raises important issues of interface ergonomics and HCI. In practice, pauses in the sonic texture gave time for shifts in spatial imaging to be prepared.

Due to the improvised nature of the performance, the spatialist has no instructions regarding compositional intent of how a sound should be spatialised. As Denis Smalley notes [19], many sounds imply space and movement anyway. The spatialist has to choose whether to embrace or challenge this, as well as ascertaining whether the different sound sources may be in conflict or unity, and how this should influence the spatialisation.

B. From the Improvisers' Perspective

The spatialist was able to determine the final presentation of the sounds to both the improvisers and the audience. Levels of ambiguity became apparent through surround spatialisation as the improvisers could not anticipate the spatial origin of the sounds. The immediacy of the spatialisation meant the improvisers were instantly enveloped by their own gestures, heightening and extending aural and spacial awareness.

Ultimately, sound spatialisation addresses issues relating to the way an improviser constructs musical meaning. If we accept listening as not merely a passive exposure to sensory phenomena, but an active process of constructing meaning, then it becomes clear that the spatial profile of a sound will affect how that meaning is constructed. Much of our listening and capacity for signification of sounds is mediated by bodily and spatial metaphors, as improviser and theorist David Borgo notes, asserting that 'Our musical vocabularies are in fact filled with embodied metaphors: pitches are high or low; sounds are close or distant; textures are dense or sparse' [20]. The spatialisation system was able to exploit this, continuously altering the timbral characteristics of the sounds, shaping the course of the improvisation as it affected the way in which the turntablist and the laptop performer listened to the sounds they produced, and the way in which the audience listened to the improvisation. Unity or conflict between the improvisers was made explicit through discrete placement and inter-manipulation of their separate stereo feeds.

VI. Recent Performances

A. Vreemdeling: A Performance with Joystick Control

Composer Robert van Heumen has recently completed a two week residency with the Resound system, using a SuperCollider patch to spatialise his stereophonic electroacoustic work *Vreemdeling* in a performance that took place at Culture Lab on 13th June 2008. Three simultaneous stereophonic layers were spatialised using a joystick controller, with SuperCollider performing the intermediate logic between the joystick and the Resound client application [21].

B. A Second Turntables and Electronics Trio

A second improvisation with turntables, electronics and spatialist was presented as part of the same performance. The same physical interface – the MIDI controller described previously – was used for spatialisation, but the configuration was rather different. More semi-automated behaviours were used following recent developments to the Resound system. This, along with the use of different materials by the sounding musicians, resulted in an altogether different dynamic during the improvisation.

VII. Future Work

Clearly there is scope for further exploration of the broad research questions posed earlier, particularly in light of the subsequent performances just described. This final section describes, in no specific order, some future considerations.

Very brief feedback has been given from the perspective of the performers. This could certainly be elaborated. Further empirical research into how multiple sources can be independently spatialised would be useful, as would a deeper analysis of the interaction between sounding musicians and the spatialist. It would be useful also to gather feedback from audience members.

From the HCI perspective, a fuller discussion of, and further experimentation with the control mapping of source-channel-to-loudspeaker combinations would clearly be beneficial as this would have a significant impact on the logistics of performance from the spatialist's perspective.

In terms of sound spatialisation as an instrument, a review of how the use of the Resound system in particular differs from other approaches to live, improvised sound spatialisation will be helpful. Specifically, a study comparing the present approach to the perhaps more common scenario in which electronic performers control their own spatialisation directly, will be worthwhile. This point will be particularly useful in comparing the improvisation trio performances – where the spatialist is an independent musician – and the performance of *Vreemdeling*, where the composer is in full control. There is clearly also a discussion surrounding the difference between fixed medium and strictly live performance.

Further exploration of the possibilities offered up by alternative control interfaces is also warranted. New interface technologies such as sensor based instruments open up the possibility of developing a control surface which offers a legibility of gesture, providing intuitive links between the movements of the spatialist and the way in which the sound is manipulated, and allowing for dexterous control of the sound sources. The use of multi-touch table-top interfaces in musical applications is already subject to investigation [22].

The audience and the other musicians may construct musical meaning through an understanding of the performative aspects of the improvisation, and perceiving a connection between the physical movements of the performer and the sounds produced, or the way in which sounds are manipulated. This semiotic dimension of movement during a performance is a common concern of musicians developing 'virtual' instruments. Suguru Goto, who developed the SuperPolm MIDI violin, refers to researcher Claude Cadoz who suggested semiotic gesture as a possible category of gesture, describing 'gestural behaviours that function to make others know: the gestures that produce an informative message destined for the environment' [23]. In developing an instrument for spatialisation, it may be desirable to dramatically relate gestures to the movement of sounds, enabling sweeping arms to craft sweeping pans, or it may be that such obvious relations would place the theatre of the movements over the manipulation of the sounds. It thus remains a significant point of interest to investigate the effectiveness of different types of control interface with the Resound system in the context of free improvisations.

References

[1] P. Bell, A. Parkinson, and J. Mooney, "Dead Dialogues: Sound Spatialisation in Free Improvisation," video online at *http://video.google.co.uk/videoplay?docid=47006225079839391*

[2] J. Mooney, "Sound Diffusion Systems for the Live Performance of Electroacoustic Music," Ph.D. thesis, University of Sheffield, 2005.

[3] J. Mooney and D. Moore, "Resound: Open-Source Multi-Loudspeaker Sound Spatialisation," *Proc. ICMC 08*, Belfast, UK, August 2008.

[4] A. Moore, D. Moore, and J. Mooney, "M2 Diffusion: The Live Diffusion of Sound in Space," *Proc. ICMC 04*, Miami, Florida, pp. 317–320, July 2004.

[5] L. Küpper, "Analysis of the Spatial Parameter: Psychoacoustic Measurements in Sound Cupolas," in F. Barriere and G. Bennett [Eds.], *Composition / Diffusion in Electroacoustic Music*, Bourges: Editions Mnemosyne, 1998, pp. 289–314.

[6] H. Tutschku, "On the Interpretation of Multi-Channel Electroacoustic Works on Loudspeaker-Orchestras: Some Thoughts on the GRM-Acousmonium and BEAST," *Journal of Electroacoustic Music*, Vol. 14, pp. 14–16, 2002.

[7] J. Harrison, "Diffusion: Theories and Practices, with Particular Reference to the BEAST System," *EContact!*, Vol. 2(4). Online at *http://cec.concordia.ca/econtact/Diffusion/Beast.htm*

[8] C. Roads, J. Kuchera-Morin, and S. Pope, "The Creatophone Sound Spatialisation Project." Online at *http://www.ccmrc.ucsb.edu/wp/SpatialSnd.2.pdf*

[9] C. Clozier, "Presentation of the Gmebaphone Concept and the Cybernephone Instrument," in F. Barriere and G. Bennett [Eds.], *Composition / Diffusion in Electroacoustic Music*, Bourges: Editions Mnemosyne, 1998, pp. 266–281.

[10] C. Rolfe, "A Practical Guide to Diffusion," *EContact!*, Vol. 2(4). Online at *http://cec.concordia.ca/econtact/Diffusion/pracdiff.htm*

[11] D. Berezan, "Flux: Live-Acousmatic Performance and Composition," *EMS 07*, Leicester, June 2007. Online at *www.novars.manchester.ac.uk/indexdocs/Flux-Berezan-EMS2007.pdf*

[12] W. Gaver, J. Beaver, and S. Benford, "Ambiguity as a Resource for Design," *Proc. CHI03*, Ft. Lauderdale. New York: ACM Press, 2003. Online at *www.equator.ac.uk/var/uploads/2002-gaver-0.pdf*

[13] S. Godlovitch, *Musical Performance: A Philosophical Study*. London: Routledge, 1998.

[14] J. Ferguson and P. Bell, "The Role of Ambiguity within Musical Creativity," *Leonardo Electronic Almanac*, published as an electronic supplement to *Leonardo Music Journal*, vol. 17, 2007. Online at *http://www.leonardo.info/lmj/lmj17supplement.html*

[15] "Auditory Environments 0803: Interdisciplinary Seminar on Sound Spatialisation and Immersive Audio." Programme online at *http://culturelab.ncl.ac.uk/auditoryenvironments/0803*

[16] J. Mooney, "Auditory Environments 0803," Flickr photo set available online at *http: // www.flickr.com / photos / jamesmooney / sets / 72157604195424010*

[17] Resound website online at *http://resound.sourceforge.net*

[18] J. Mooney and D. Moore, "A Concept-Based Model for the Live Diffusion of Sound via Multiple Loudspeakers," *Proc. DMRN 07*, Leeds, UK, 2007. Online at *http://www.james-mooney.co.uk/publications*

[19] D. Smalley, "Spectromorphology: explaining sound-shapes", Organised Sound 2: 107-26, Cambridge University Press, 1997.

[20] D. Borgo, *Sync or Swarm: Improvising Music in a Complex Age*, New York, Continuum, 2007.

[21] R. van Heumen, "Vreemdeling," blog online at *http://hardhatarea.com/vreemdeling*.

[22] S. Jordà, G. Geiger, M. Alonso, and M. Kaltenbrunner, "The reacTable: Exploring the Synergy between Live Music Performance and Tabletop Tangible Interface," *Proc. TEI 07*, Baton Rouge, Louisiana, 2007.

[23] S. Goto, The Aesthetics and Technological Aspects of Virtual Musical Instruments, Leonardo Music Journal, Vol 9, 1999.

Traditional and digital music instruments : a relationship based on a interactive model

Paulo Ferreira-Lopes
CITAR Center for Science and Technology in Arts
Portuguese Catholic University
ZKM|Zentrum für Kunst und Medientechnologie - Institut für Musik Akustik
pfl@zkm.de

Abstract – **In the present work some aspects of the influence of the digital music instruments on composition methods are observed. Some consequences of the relationship between traditional instruments and digital music instruments result on a triangular interactive process. As an analytical approach to understand this relationship and the association process between traditional instruments and digital music instruments, a typology of interaction for the musical performance based on this instrumental configuration is proposed. The deduced classes are based upon the observation and systematization of my work as a composer. The proposed model aims to contribute towards an unifying terminology and systematization of some of the major questions that arise from the coexistence between two different paradigms (traditional and digital music instruments) in the universe of live electronic music.**

I. Building a systematization method

My activity as composer implies a constant research, in the field of sound materials, namely timbre. Although this research activity does not always occur as a systematic strategy, I am very focused on understanding and developing some interaction models on the relationship between composition and interpretation. These models, adapted to precise situations, make it possible to establish, *a priori*, a kind of memory catalogue in order to reflect a global paradigm, within a strictly musical point of view, where composer, composition, musicians and instruments interact in a macro scale [1]. In order to explain the connection modes and communication models, within the exchange and cooperative relations among traditional musical instruments and digital music instruments we propose a paradigm based on a typology form. Traditional music instruments and digital music instruments are able to influence and interact mutually through the actions of performers. Consequently, in my work, the interaction classification in the form of a typology results from the need to systematize some categories and hierarchical relations in this universe. A global approach to these problems leads to two main questions:

- the systematization categories derived from the relations between musicians and instruments;
- the adaptation and understanding of the musical resources and techniques to the technological contexts - better technological choices or improved technical solutions - in order to realize a musical work.

II. Study case

Through the interpretation of a live electronic music work, the relations established between all instruments able to operate and play in real time (digital music instruments and traditional music instruments) - shows us different characteristics and qualities [6]. These differences can change according to the technological devices or the technological infrastructure associated to the realization of the work [3], such as:

- traditional instrument and music on support (CD or tape);
- traditional instrument and digital music instrument with limited access (digital MIDI sampler, digital MIDI synthesizer, effects processor, etc.);
- traditional instrument and digital music instrument based on the computer (open programming access).

Therefore, the technological device category enables the definition of the kind of interactions limits between the musicians and the instruments as well as the interpretable musical parameters. For example, in a work for traditional instruments and music fixed on support (Tape or CD), it is very difficult to interpret the time, because time and durations associated to the recorded musical contents on support are unchangeable [7]. On the other hand, in a work for traditional instruments and digital music instruments (able to operate in real time) as well as in a work for traditional instruments, the interpretation of time has a very large space of variation, operated with much more flexibility. Thus, we perceive that the characteristics and qualities of the relations between traditional instruments and digital music instruments can be classified according to interaction categories. In the last few years, we have developed specific strategies based on our experiments to conceive the right communication approach between traditional instruments and digital instruments. In this framework we have introduced typological descriptors aiming to characterize the interactive processes employed in our works. It is our belief that the typology of an interactive process is delimited by two principal categories: *kind* and *directionality* (see Fig. 1). As far as *kind* is concerned, we divide it into two subcategories:

- discrete;
- continuous.

Regarding *directionality*, we also divide it into two subcategories:

- one-way;
- two-ways.

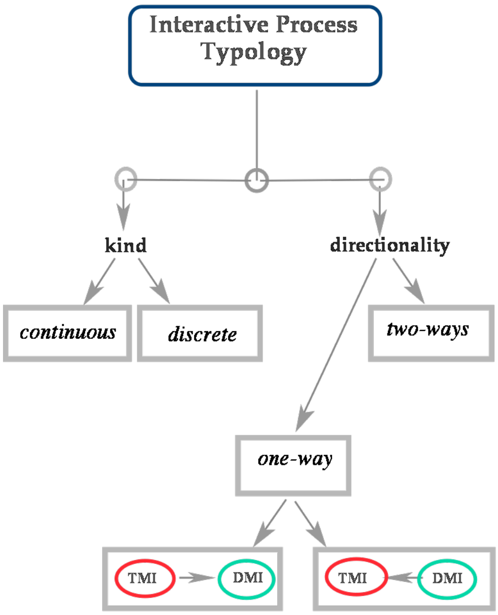


Fig. 1. Typology of interactive processes.

Regarding *directionality*, in the case of a one-way process, it can be further divided into two subcategories:

- when the traditional music instrument is able to influence the musical results of the digital music instrument (TMI---> DMI);
- when the digital music instrument is able to influence the musical results of the traditional music instruments; (TMI < --- DMI).

In the Table 1 we can see clearly the different regions produced by the division categories.

TABLE I.
INTERACTION CATEGORIES TYPOLOGIES INTERSECTION

		DIRECTIONALITY		
		One-way		**Two-ways**
		T>D	D>T	-
KIND	**Discrete**	TDD1	DTD1	TWDI
	Continuous	TDC1	DTC1	TWCI

III. THE CATEGORIES DESCRIPTION

3.1. The Kind Category : Discrete Interaction

We begin this chapter whit an example of a discrete interactive process : when a traditional music instrument controls the sound produced by a digital music instrument through procedures controlled with impulses. It can be the situation where a note, a frequency or one dynamic peak from the traditional musical instrument signals starts: categories

- a sound or a whole collection of pre-recorded sounds;
- a global effect (reverberation, delay, etc...).

The principal characteristic of a discrete interactive process, being it one-way or bidirectional, can be reduced to an impression of *globality*. In this case, the notion of *globality*, even if able to express different meanings, is circumscribed by us to the aspects that refer to a real time transformation of the sound matter [10]. *The globality*, as the main characteristic of this process, is a consequence of the control mode. In fact all that is related to the handling and the sound processing is operated from a single event, where data control continuity does not exist. In this network, each point represents a place of intersection where a value, a data or a single impulse, determine the beginning or the end of an event, a treatment or a musical transformation. In this type of structure, the communication between man and instruments, is realized, in a general way, through very distant coordinates (space and temporality):

- where the events cannot establish, successively or alternatively, a relation between them;
- where an event cannot be transformed in a continuous mode (from the starting point up to the end point);
- where the existence of an transitional space, between the temporal occurrence of each point cannot establish a kind of relational memory between the performance and the musical results of the electronic part.

Even if this structuring method appears very effective, from the point of view of system stability like the control of the multiple digital music instruments, it limits the continuity of musical gesture [2]. This aspect reduces drastically the interpretation potentialities of the electronic part in a musical work.

3.2. The Kind Category: Continuous Interaction

The most important characteristic of a continuous interaction process consists on the distribution of interventions in which each intervention usually relates to a specific process [11]. These interventions are more commonly materialized in the form of sound processing or of musical material generation. This means that: the increase of a continuous interaction process between the musicians and the digital music instruments imposes specific procedures to each situation [9]. This principle presupposes one permanent continuous data exchange during the realization of live electronic music work and implies that all the parameters, controls and adjustments are carried out in real time, both in the level of musical generation and of sound processing. One example of such a continuous interactive process: when a traditional music instrument controls the digital musical instrument sound results through a continuous data exchange. In this method an additional level is required in order to accomplish the acquisition and the data representation extracted from the controller instrument (in this case a traditional music instrument). Generally, the real time sound processing or a sound generating control method depends from the data tracking method catches, during the musical performance. A continuous interactive process requires a continuous data tracking method in order to represent a continuous flow of numerical values. The assembly of these data generally entails two successive steps:

- the step related to data acquisition;
- the step related to the previous gathered data representation.

The step concerning data acquisition can be carried out in three forms:

- data acquisition, based on the audio signal analysis coming from the traditional instrument;
- data acquisition, based on movements, mechanical actions of pressure of displacement tracking through various sensor types;
- data acquisition, based on the combination of the above.

The main purpose of the representation step is to adapt and transcribe the data provided by the analysis from the signal, the sensors or both. The greatest difficulties regarding data transcription lie on the adaptations of common variables to a model tailored to a specific concretization [12]. In this type of structuring, the communication between the man and the instruments allows that:

- various events can successively or alternatively establish a relation between them;
- an event can be transformed in a continuous mode (from the starting point up to the end point);
- a transitional space, between the temporal occurrence of each point establishes a kind of relational memory between the performance and the musical results of the electronic part.

3.3. The Directionality Category : One-way interaction

As was mentioned in the introduction (see also Fig. 1 and Table 1) the directions in which an interactive process progress can be taken in different ways. The advantage of identifying the direction among which the process progresses, is that it is possible to quantify the influence between the entities involved on the process. On the other hand, understanding the direction among which an interactive process progresses also makes it possible to deduce three aspects simultaneously:

- the number of entities engaged in the process and their principal qualities;
- the hierarchical placement of each entity;
- the entities which act and those which are affected by the actions.

Taking the literal significance of the interaction concept in the context of a communication process, one of the first meanings, and the most spontaneous one, establishes a significance link to a kind of reciprocal game between two or several entities [8]. This concept of exchange and reciprocal influence in an interactive process characterizes the different entities related to the process as human beings. However, according to Jensen [5], one can deduce that the medias also have some singular qualities and characteristics, which are able to influence the behaviours and actions of human beings:

> *"... the concept of interaction is a concept directly correlated with the communication, which implies among the process of exchange a permeability to the medias characteristics ...." [5].*

The one-way interaction typology, such as we discuss it in the musical context, represents the particular case of the interaction because the communication channel works in one way only. In this case, the process evolution vectors are spread in only one direction. This situation implies that the entity affected by these vectors is not able to reply because the communication structure does not contain a return channel. In the case where the interaction process is carried out in a one-way direction, we state two distinct situations:

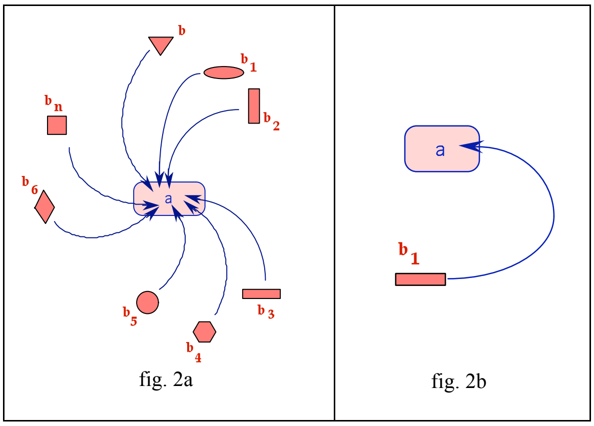


fig. 2a

fig. 2b

- one or more performers influence, through their instruments or through external controllers (see fig. 2a and 2b) the musical results of the digital music instruments;

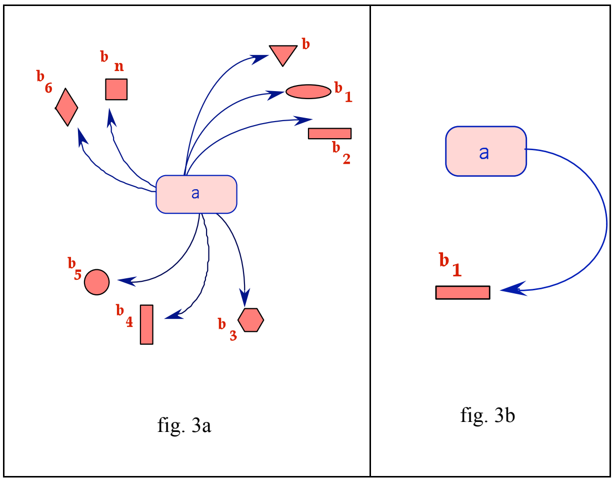


fig. 3a

fig. 3b

- one or more digital music instruments interact and influence the musical results of one or several interpreters (see fig. 3a and fig. 3b).

3.4. The Directionality Category : Two-way interaction

In the cases where interactive typologies are based on a two-way communication method, the principles of this typology can be considered as a duplication of the communication chain between the entities related to the process, but in opposite directions. In this way, we can conceive a general category of bidirectional interaction, where it is possible to include at least two typologies of interaction. The first typology: we can observe a similar preponderance of influence between each instrument. In this typology, it is possible to observe a process in which the traditional music instrument controls the musical results of the digital music instrument and at the same time the processes of the control - in the form of transformations into real time - that the digital music instrument operates on the achievements of the traditional musical instrument. We can illustrate this typology through the following example: data extracted from the dynamics analysis of a traditional music instrument make it possible to start the sound or music sequences previously sampled and stored on hard disk. In tandem with this, the digital music instrument transforms the sound of the traditional music instrument in the form of global effects: reverberations effect, harmonizer effect, etc... In this example, we mentioned the effects operated on the traditional music instrument as global features, however the transformations operated by the digital music instrument could have been more complex. It only depends on the choice of typological kind (continuous or discrete) and consequently in the mapping complexity of control parameters between the physical action of a human (TMI) and a programmed environment within the computer (DMI) [4].

IV. Conclusion

The background upon which we characterize the interactive process between a traditional music instrument and a digital music instrument is based upon a typology based on two main categories: *kind* and *directionality*. This means that, on any interactive process, there is one stratification related to *kind* and another one related to *directionality*.

Hence, *kind* reveals the internal type of interaction process articulation (discrete, continuous), and *directionality* dictates which entities (traditional or digital music instruments) will be the object of the process assignments.

In the situations where it is intended that the digital music instrument achieves highly complex musical results and a “humanized” performance, we notice that what results from sound processing only denotes the desired complexity if the process of articulation between the traditional and digital music instruments is accomplished by a kind of continuous communication. Within this framework we also remark that the quality of a “humanized” performance implies that the traditional musical instrument assumes a non-common category: the interface controller of the digital music instrument.

In cases where the main accomplishment intended for the interaction is a global sound processing effect (reverberation, harmonizer, chorus) applied to the traditional musical instrument through the digital music instrument, we observe two major problems:

- first, the computer cannot determine beforehand what the musician will carry out in the performance due to the fact that the computer is used simply as a transformer entity and the communication chain had only a one-way channel;
- second, the magnitude of the sound resolution scales and the sound fragmentation dimensions do not allow a micro local manipulation and micro temporal isolation of the different layers from the sound spectrum. This type of constraint implies that the sound manipulation and the spectral transformations can only be operated on undifferentiated and global sound mass.

The second typology is based on situations where it becomes possible to establish two channels of communication, a two-way interaction, the traditional musical instrument can have a dual function: as transformed and as transformer. In this kind of situation, it is possible simultaneously:

- through a channel, to configure an interaction process, favouring digital musical instrument operations on the traditional instrument ;
- through the second communication channel, it is possible that the traditional music instrument makes the assignment to the transformations settings itself, that will be produced on its own sound.

V. Acknowledgments

This article is part of the theoretical research from the project PRICES supported by CITAR – Research Center for Science and Technology in Arts from the Portuguese Catholic University and the FCT - Foundation for Science and Technology , Portugal (POCI 2010). Special thanks to António de Sousa Dias, Daniela Coimbra and Álvaro Barbosa for reviewing this work.

VI. References

[1] Ferreira-Lopes, P.; *Étude de modèles interactifs et d'interfaces de contrôle en temps réel pour la composition musicale.* Thèse de Doctorat ; Paris ; Université de Saint Denis - Paris VIII - Dép. de Sciences et Technologies des Arts, 2004

[2] Ferreira-Lopes, P., Coimbra D. and Sousa Dias, A.: "Music and Interaction: Consequences, Mutations and Metaphors of the Digital Music Instrument" *in ACTAS do 2º Workshop Luso-Galaico de Artes Digitais* ; Vila Nova Cerveira / Portugal ; 2005.

[3] Garnett, G. E.*;* The Aesthetics of Computer Music in *Computer Music Journal vol 25 nr. 1*; Massachusetts: éd. The MIT Press, 2001.

[4] Goebel, J.; "The Art of Interfacing: Senses, Sense and the Discipline of Playing Interfaces" *in The Sciences of the interfaces* (p. 306-314); Tuebingen: éd. Genista VERLAG, 1999.

[5] Jensen, J.*;* Interactivity; Tracking a new concept in media and communication Studies in *Computer Media and Communicatio* ; Oxford : Oxford University Press , 1999.

[6] Lèvy, F. ; Texte de présentation du CD-Rom "Le temps réel en musique" - *A l'occasion de l'exposition "Le temps, vite" présentée au Centre Georges Pompidou,* Paris : éd. IRCAM, 2000.

[7] Manoury, P.; "De l'incidence des systèmes en temps réels sur la création musicale" *in Actes de la Conférence ARTE E TECNOLOGIA*; Lisbonne : éd. Fondation Calouste Gulbenkian;1987.

[8] Rafaeli, S.:Interactivity : From New Media to Communication *in Sage Annual Review of Communication Research* ; Beverley Hills - Newbury Park : éd. Sage; 1988. (pp110-134)

[9] Risset, J. C.; Evolution des outils de création sonore *in Interfaces homme-machine et création musicale* ; Paris : éd. Hermes; 1999.

[10] Vaggione , H. : "L'espace composable sur quelques catégories opératoires dans la musique électroacoustique" *in M. Solomos et J-M Chouvel (Ed.): Espace : musique, philosophie.* ; Paris, L'Harmatan; 1998.

[11] Weibel, P.; "The Art of Interface Technology" *in The Sciences of the interfaces* (p. 272-281); Tuebingen : éd. Genista VERLAG ; 1999 ;

[12] Winograd, T.: Interaction Spaces for 21st Century Computing, in John Carroll (ed.), *Human-Computer Interaction in the New Millennium*, Addison-Wesley, 2001.

Memory Space

Simon Emmerson
Music, Technology and Innovation Research Centre
De Montfort University, Leicester UK
s.emmerson@dmu.ac.uk

Memory may be mapped metaphorically onto space, as in the mediaeval and renaissance *Memory Theatre* (see Frances Yates, *Art of Memory*, 1966/1992 [Reference 4]). But we now have the power to project this literally in sound in sound installation systems such as the *Klangdom*. In *Resonances* (8 channel acousmatic work, commissioned by the IMEB (Bourges) in 2007), I explore my memories of the modernist repertoire (1820-1940) using small timbral 'instants', extended, layered and spatialised. Some juxta- and superpositions are logical, others unlikely – but that is the joy of memory and creativity. But memories also fade and die ... This paper examines this work, and how the memory and spatial relationships are articulated through the material. It also presents plans for a more elaborate work to be realised in 2008-2009. In this the fixed nature of the previous work will give way to an 'evolving' acousmatic piece which changes at each performance as new spatial layers are added, others fade. The paper will be illustrated with music examples..

I. MEMORY AND SPACE

Space is not simply a geometric 'thing out there'. We are born with sight and sound, touch, taste and smell ready to initiate our particular construction of it. Space would not be perceptible without objects, textures, sounds. The sonic arts have tended to separate out taste and smell (although they have crept back in in recent more open social musical spaces) – and touch, too, unless you are a performer.

But *memory* is also spatial in two senses. Neuroscience is slowly unlocking the secrets of the most complex system observable by us – the human brain. But the nature of memory within the brain is not much understood – except that the questions are becoming more sophisticated and it is evidently distributed in many locations. But there is also a deeper link which has been exploited over some thousands of years – most extensively before writing (and more specifically printing) allowed us a short cut.

This is best described in classical, mediaeval and renaissance practices of mapping places, images and other objects of memory onto an imaginary stage in the mind – the so-called *Memory Theatre*. This was most especially examined in Frances Yates's book *The Art of Memory* (1966/1992 [Reference 4]). Starting from ideas of rhetoric inherited directly from the Greeks through Roman sources we start from the idea that natural memory can be improved or augmented through the exercise of 'artificial memory'. This is created from *places* and *images*. A place (*locus*) is easily memorised – a construction, a characteristic *location*. Images are 'forms, marks or simulacra of what we wish to remember' [Reference 4: p.22]. There have been developed rules for places and rules for images. The *loci* relate to each other such they can be walked through in the imagination and even a particular building is to be seen as the best recepticle for the totality of the loci. Each fifth locus is given a particular mark in its characteristic. The building should be empty since crowds might distract! This allows us to construct two kinds of artificial memory - *memory for things* and *memory for words*. 'Things' are not objects in the contemporary sense but can include the subjects of speech - 'the ideas we are trying to express' - while words are simply (but importantly) a vehicle to convey that and often have to be memorised in the 'correct' order. In brief, the art of artificial memory lies in the direct association of image with locus and the ability to recall one through 'visiting' the other. By the Renaissance period the building within which the memory locations were found was very often constructed as a kind of theatre with five doors, five columns and other easily memorised characteristics. Yates discusses many of these examples and their complex historical interconnections – including highly dangerous rivalries, accusations of magic, heresy and the like – such as those of Giordano Bruno (late 16th C) and Robert Fludd (early 17th C).

II. *TIME PAST*

I have for a long time been interested in exploring through composition how memory and music interact. In the 1980s I completed a series of works (*Time Past I-IV* for various solo instruments or voice and electronics) based on different 'layers' of memory from mythic and cultural, to poetic and 'technical'.

In *Time Past I* the tape-delay system of Stockhausen's *Solo* was used to create polyphony. The tape delay systems of those years had a physical spatiality now lost to the digital world. The longer the tape, the longer the time. Delay times from 7.5s to 45s – that is from the edge of short term memory into the longer term – were used. The choices for the live montage included when (and hence what) to record. Which delay playback heads were used, and whether the signal from these was fed back and re-recorded. The resulting polyphony managed to avoid the obviously canonic (so easily done with delay lines of fixed duiration). In *Time Past II* (instrumental and electronic) and *III* (acousmatic) issues of mythic time ('before recorded time') were modelled on Simha Arom's work in the Central African Republic. Most specifically the myth of the origin of music from the NgBaka, in which a hunting trap becomes a music bow and thus music is discovered. Common usage of the bow for hunting and music is found in many cultures.

In *Time Past IV* (soprano and tape), Shakespeare's Sonnet 30 ('When to the seesions of sweet silent thought, I summon up remembrance of things past ...'), the text is fragmented and then slowly reassembled as the work progresses. In a kind of archaeological reversal of time the sonnet is pieced together; fragments change order until the whole is finally revealed. A necessary struggle to recall and reconstitute. (Yates discusses Shakespeare's Globe Theatre in detail as relating to a real manifestation of Fludd's memory theatre [Reference 4: chapter 16].)

The entire sequence was influenced also by my reading Marcel Proust's *A la recherche du temps perdu* at that time. A taste, a smell, a felt texture can trigger the most profound mental shifts into memory. These are not mere recollections of the past but shifts in mental space for a brief moment 'back to where you were then'. As the 'author' remarks in final novel of the sequence, *Time Regained*:

> "… for in order to get nearer to the sound of the bell and to hear it better it was into my own depths that I had to re-descend. And this could only be because its peal had always been there, inside me, and not this sound only but also, between that distant moment and the present one, unrolled in all its vast length, the whole of that past which I was not aware that I carried about within me." [Reference [2]: p.1105]

III. THE MOMENT (EXTENDED)

Duration has no meaning in spatialised memory – but that gives its recall and realisation great flexibility. That is in externalising a memory we may subject it to all kinds of time manipulations. Historical time can be 'time stretched'. Karel Goeyvaerts talked in his early correspondence with Stockhausen (1951-53) of 'dead sounds' - effectively of virtually no duration [3] - the impossibility of which did not stop the idea having resonance in Stockhausen's later thinking. The instantaneous and the eternal meet in William Blake's famous dictum: 'He who kisses the joy as it flies lives in eternity's sunrise' which was of course the motto for Stockhausen's *Momente*. That same composer's *Hymnen* had space built in as we 'flew' with him across continents, glimpsing national anthem groups and regions, finally finishing up in the Utopia of 'pluramon' a striking image of both memory and prediction. In fact the anthems of *Hymnen* are increasingly the objects of memory as they fade, change meaning and simply disappear. The Cold War was once the dominant contemporary world-image and is now faded. The metaphors of space and time blur inevitably in listening to this work. This explains the fragility of the postmodern idea of 'double coding' – we increasingly fail to remember the 'references'. And, furthermore, each of us has a unique pattern of memory fades – we can call this a unique resonance.

IV. SPACE AND TIME

Sound projection (on such as the acousmonium) allows us to play with sound in space in performance. But the advent of multi-channel surround space allows composers to control that to an even greater extent – to control more clearly a spatial polyphony and to allow sounds to be heard 'da lontano' at the same time as an intimate close soundfield is also present.

György Ligeti played with the spatiality of memory in his orchestral work *Lontano*. He brilliantly manipulates the tonal space of harmony into an allusion of real space and distance. This in turn reminds us of things we think we have remembered –

> "… the very sound of horns has a 'historical perspective'. … an allusion, a reference to certain elements of late romantic music. … particularly of Bruckner and Mahler, but also of Wagner. … Well, there are many similar – not quotations, but allusions in *Lontano*. I would say that as well as spatial distance, there is also temporal distance; that is to say, we can grasp the work only within our tradition, within a certain musical education. … it does not treat exact quotations from late romantic music, but certain types of late romantic music are just touched upon. … the temporal distancing evokes also a spatial distancing. The horns can be heard from a distance and from long ago …" [Reference 1: p.93]

This work (its soundworld and its ideas) undoubtedly influenced my own recent work *Resonances*.

V. *RESONANCES*

Most recently (2007) I completed *Resonances* (8-channel acousmatic). This was commissioned by the Institut International de Musique Electroacoustique de Bourges and was first performed at the Festival Synthèse, Bourges in June 2007.

Resonances 'plays with' our memories of the timbres and colours of the modernist repertoire (with considerable reference also to some 19th C precursors from which it is said to have evolved) (1820-1940). I always loved the music of the 20th C having come to it earlier than more classical or romantic repertoires (the Bartok string quartets were my earliest record purchases). In the great tradition of British education I was asked to teach a course on modernism in music at short notice and relatively unqualified; but I realised in preparing such a programme how much I already knew by ear and how little I knew of the background theory. The sound came first and I have always attempted to teach with this priority – the experience (and memory) of the soundworld first followed by its construction and historical context in tandem.

In this work we hear glimpses of music we think we know; no melody or rhythm, just an instant of 'colour', frozen and moving – 'eternity in a moment'. Colours may be vivid and clear, or sometimes dark and opaque. This is a personal choice from my memories of a repertoire I love but which may not last another century. Memory changes perspectives, transforms and shuffles, creating unique resonances in each of us.

VI. SPATIALISATION OF TIMBRAL MEMORY IN *RESONANCES*

I will now describe the *spatialisation of timbral memory* in *Resonances* - how layers and movements of sound are not only spatialised for 'acousmatic' reasons but as an analogy to spatial thinking about history and memory. Personally I 'see' time and space in front of me while composing in the studio. While using 'classical' aural approaches to sound combination – that is with no imposed theory on the one hand or overt narrative on the other – I also found that the 'memory time' relations started to speak directly to me, indeed intuition and free association often indicate a subconscious hard at work. But there was one explicit decision - the opening. As Adorno said, it all began with late Beethoven. The suspended discord at the centre of the fourth movement of the Ninth Symphony is first and foremost an extraordinary sound in which harmony dissolves into timbre. This is where we begin. This in turn dissolves into that great enigma of the mid-19th century, the 'Tristan chord'. Do I claim that decision is based purely on the sound? I'm not so sure ... but I now 'played' with another version of the chord (in Debussy's *Prélude à 'L'après-midi d'un faun'*) which my memory has suggested to me. This collapse and expansion are at once timbral and spatial, they literally grow around the listener.

But in addition to the timbral transformations that articulate this juxtaposition, the history of 19th C music is also spatialised around the listener. This is further reinforced through the metaphor of the river. There are allusions to the Rhine (although it has moved slightly – another Debussy-Wagner constellation whose sketch was titled *La cathédrale engloutie ... sous le Rhin*) and later in the piece a Swan boat that sails to another country and century (Wagner to Ligeti).

Memory shuffles both time and space: this is often described through the metaphor of the window 'through' to another world displaced from that in the current foreground. But in turn that foreground can curl back and retreat to the distance. We glimpse times yet to come, new sounds and timbres. Sometimes the space of another planet (as in Schoenberg's imagination). In addition to the window there is the mirror – the perfect space-time construction of modernism (especially for the Webernians). A purely spatial construct is elevated to a major metaphor of time reversal – 'in my end is my beginning'. So Schoenberg and Webern are found transformed in space and time as each point is placed in the memory map.

But in memory is also forgetfulness and ... we forgot the Russians! So there is a burst of colour and we go from the slow movement of a coronation to the ecstatic fury of a dance to the death – where modernism left the body for the best part of the 20th century, and where the piece concludes.

VII. FUTURE PLANS

I am planning a work which explores these relationships in a more dynamic and open work. The *Klangdom* (or its equivalent installation) can become a literal representation of such a spatialisation of memory in sound. Software would control movements of the different layers of 'memory' (and their eventual fading). While conceived as a primarily acousmatic work there will be a 'live mix element' resulting in similar but evolving performances over time with each performance adding a new layer while other memories transform and eventually fade out. It is planned for completion in 2009.

REFERENCES

[1] Ligeti, György (with Péter Várnai, Josef Häusler, Claude Samuel), *György Ligeti in Conversation*, London: Eulenburg Books, 1983

[2] Proust, Marcel, *Remembrance of Things Past* (3 vols), tr. C.K. Scott Moncrieff and Terence Kilmartin, Harmondsworth: Penguin Books, 1983

[3] Toop, Richard, 'Stockhausen and the Sine-Wave: The Story of an Ambiguous Relationship', *Musical Quarterly*, LXV (July 1979), pp. 379-391.

[4] Yates, Frances, *The Art of Memory*, London: Pimlico, 1992 (original, London: Routledge & Kegan Paul, 1966)

Echi tra le Volte, a sound design project for churches

Andrea Taroppi
Conservatorio di musica di Como, Italy,
tesi biennio specialistico in Musica Elettronica e Tecnologie del Suono
taroppi@infinito.it

Abstract — **This work is about the relation between music and architecture. In particular we are interested in the concept of space, as the way where music and architecture meet each other. The study of this topic offer the starting point to the development of *Echi tra le Volte*, a music installation for churches, where sounds are from the natural reverb of the place, excited by a sinusoidal impulses, which receive its pitches from a genetic algorithm**

I.Introduction

The aim of the *Echi tra le Volte* project is to create a music which could be closely related to a specific building. Such a music would be the acoustic completion of the visual aspect when visiting valuable buildings.

When we think about music composed to be site specific, music with an intense link to a specific site, either a building or a natural place, we can observe that, taking due account of differences, generally speaking the used strategy is a descriptive or analogical one.

Such a methodology is based on the common use of numbers and proportion, for example the golden ratio, in music as well as in architecture.

Roots of this equivalence are in ancient Greek, when Pitagora found a relation between music perception and physical world, discovering the law that links string length with the pitch of the sound produced.

Consequently, it started and could develop an aesthetic of the number, well summarized in the words of the Italian architect Leon Battista Alberti (1404 – 1472), who wrote in his *De re aedificatoria*: “Those numbers that have the power to give smartness to sounds, that is so pleasant to our ear, are the same that could fill our eyes and our soul with admirable delight [1].”

Roughly in those same years, it is considered that Guillaume Dufay used elements related to measures and size of the Florence Cathedral “Santa Maria del Fiore”, for the composition of his motet *Nuper rosarum flores*. According to some studies, e.g. in [2], the dimensional proportions of the church, to witch the music is intended are related to the music structure of the motet’s tenor and contratenor.

Also, there are examples of inverse path: from music to architecture. One for all the Steven Holl’s *Stretto house*, “designed as a parallel to Béla Bartok’s *Music for Strings, Percussion and Celesta* [3].”

Finally, showing that the idea and desire to link eyesight and hearing could be found also outside the Western culture, should be mentioned the case of the Australian aboriginals and the songs they use to determine boundaries and particular areas. In those songs “The contour of the melody […] describes the contour of the land with which it is associated [4].”

In all cited cases, the link between sound and visual is in the geometric features, or more generally in quantifiable ones, as elements by which to make the translation from architecture to music.

However, it is not said that what is clear in terms of interpretation, it is so in terms of perception as well.

Also, it should be noted, that we are in front of a one way path: from architecture to music *or* from music to architecture. But It could be better: from architecture to music *and* from music to architecture; as a sort of dialogue between the two areas, so that not only the building influences the perception of music, but also the music influences the perception of the building.

So I think it is not in numbers, measures or proportions we can find the better channel of communication between sound and site, or focusing on architecture, between a music and a building. Furthermore, a numbers based relationship could not be considered peculiar of these two disciplines. Indeed, also in other areas, such as painting or even in literature, we can find measures and proportions.

Rather, it is in space that happens the meeting between music and architecture, because space is the common place of existence of music and architecture, and the object of design not only for architecture but for music too.

II.Music as Space Designer

Ambient perception is modified by sounds, in at least two ways. One relates to where sounds are perceived, that does not necessarily coincide with the position of sound sources. Think about the “in the middle” position of the singer, i.e. as if the voice come from a point on the wall, exactly halfway between the loudspeakers; a feeling we all can test, listening to almost all the popular music by a stereo equipment.

In addition, sounds can affect on the perception of a particular place, acting on aesthetical or emotional level.

This happens, when sounds contributes even heavily to make more or less pleasant a place. Think to the background music, or muzak, in supermarkets, waiting-rooms and so on.

We can read the most famous silent piece by John Cage *4’33”*, as a case in which space gives birth to a relation between music and the place it is played in, without use of any analogue similarity or quantifiable element.

In 4’33” playing instrument is not a traditional one, but the concert room itself. And all the sounds in it are no longer noises, but *the* composition. Soundscape becomes at the same time material for and result of composition, in a no end feedback.

In *Modes of Interference* by the Italian Agostino di Scipio, the composer explores the Larsen effect by the feedback from a microphone – loudspeaker system, so that “the room acoustics does not simply *host* the performance, but shapes it and contributes actively to it, while also setting precise material conditions and boundaries for it to happen [5].”

Something similar happened in the XVI century Venice, where composers like Andrea Gabrieli were able to make a composition purpose use of the San Marco Cathedral reverb.

In another Di Scipio’s work: *Interactive Island (Sea Lights and Colors)*, we can find this kind of dialectical exchange between sounds and ambience. In *Interactive Island,* all sounds in the performance place are “captured by a number of microphones and sent back to the computer. The computer analyzes the numerical difference between the waveform of the feedback signal and that of the synthesis signal. The difference tracks the timbre modifications resulting from amplification and room acoustics [6].” And gives to the computer, values to be used as control signals for the automated composition process.

So, “both the “dead,” abstract data structure captured in the computer and the live room acoustics of the material ambience become responsible for the development of the musical flow (*ibidem*).”

In conclusion we can observe that, differently from a numerical kind approach, in a space oriented relation between music and architecture, it seems pre-eminent the will to achieve an exchange of information, resulting in a more site specific music.

III.Echi tra le Volte

From this point of view, starts the *Echi tra le Volte* (Italian for: echoes among vaults) project.

The here presented version was commissioned to be realized for churches and was performed from 5 to 25 in November 2007, at the "Santa Maria del Popolo" church in Vigevano (Italy)

The work is in two phases: the pitches generation, and the execution. In the first one, a genetic algorithm generates sequences of numbers, representative intervals from a base pitch. In the second phase, after defined the base pitch from the dimension of the church, a patch of the software "pure data" ([PD] http://www-crca.ucsd.edu/~msp/index.htm) uses the values from the first phase to set the pitches of a sinusoidal grain generator. The output from the generator sent to a four channels loudspeakers system, works as an exciter of the natural reverb of the church.

IV.Generation Process

A library named GA for the genetic algorithm use in musical composition, created before [7] for Open Music (http://recherche.ircam.fr/equipes/repmus/OpenMusic), was useful for getting the values of the generation process.

I obtained 8 evolution steps, using fitness functions derived by the analysis of interval classes in Gregorian chant [8].

The evolution process goes from randomness to order, generating sequences developed around a *corda di recita* (reciting tone) avoiding tritons and others prohibited intervals, or too wide ones. A specific kind of mutation operator was created for the generation of neuma like groups of values.

After reaching the maximal complexity, slowly the process comes back again to a chaotic situation, allowing a new generation to start again**.**

V.Synthesis And Realization

The getting values are loaded on PD and after being converted in hertz, sent to a sinusoidal oscillator.

A train of sinusoidal impulses is created multiplying the output of the oscillator by the output of a continuous table reader (as in fig 1).

There are 6 table to choose from, everyone with a different curve stored: gaussian, hanning, percussion envelope etc. Every impulse have a stochastically variable duration between 40 and 60 millisecond, while the inter-offset time is between 20 and 30 millisecond. After a time between 1 and 3 seconds the table reader stops and the train impulse (one note effect) comes to its end.

Those values are derived by the reverb time and the dimension measures of the church.

For example, it was measured the reverb time in 4 different point of the church, and the average value was 3 seconds. This value was used as the maximum duration time of the train impulses.

At the end, a specific algorithm gives every train impulse its amplitude and an increasing type envelope. So every value from the generation process results in a note of a virtual Gregorian chant. The time sequence of the notes follows the different evolutionary steps described before: from a medium of one note per 30 seconds, at the beginning in the random phase of the generation process, to a progressive approaching of little groups of note as the stochastical appearance of neumas, until a more fluent output, in imitation of long melismas.

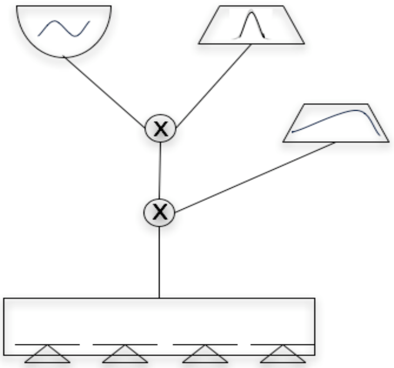


Fig. 1. PD patch scheme**.**
Impulses are created multiplying the output of the oscillator (top left) by the output of the continuous table reader (top right), another table gives the envelope (centre). Then, the four speakers spatialization system.

Then, again, notes leave each others progressively, and the system comes back to the rarefied starting point situation.

A computer with the PD patch was installed in the church and connected to a four loudspeakers system.

This spatialization system is organized so that every grain or impulse is sent to an always different couple of loudspeakers. The two loudspeakers are in phase opposition, simply multiplying by −1 one of the two outputs.

This version of *Echi tra le Volte* completes all the 8 steps in about 7 hours.

For technical reason it was no possible to change the values from the Open Music generation process, loaded in PD; so after that time the process start again with the same values.

However, this is not a big trouble because others parameters are always different, so it is like to read a score every time altering everything but the note pitches. Indeed, it was possible to let *Echi tra le Volte* goes on for about 32 hours (from Friday morning to Saturday evening) without interruption, also in the night (during closing time!), and according to the visitors opinions the result was good.

VI.CONCLUSIONS

Echi tra le Volte makes possible the direct participation of the building *on* the sound and musical structure, extending the sound synthesis chain outside the computer into the performance space.

This is the main difference from works like Di Scipio's *Interactive Island,* where computer elaboration is an answer *to* and *after* the action (proposal) of the performance place.

Besides, the *Echi tra le Volte* use of elements derived from Gregorian chant, it seems to add one more dimension in the achievement of a full relationship between music and architecture.

The result is a composition strongly linked with the building, so that it will play completely different in others places. In this sense, It could be say that *Echi tra le Volte* is the sound of the specific hosting architecture.

The project gives the occasion to think over the relation between music and architecture as a synergical connection, a way of linking sound and vision where the two (music and architecture as well as sound and visual) affect each others.

AKNOWLEDGMENT

I wish to thank M° Giovanni Cospito of the Electronic Music department in the "Conservatorio di Como", The director of the Conservatorio M° Luca Bassetto, all the Electronic Music department and the other teachers and various people who I met during those studying years in Como.

Thanks to the "Centro Ricerche per l'Ecologia Acustica" and "Le Zolle" for any help during the execution days of *Echi tra le volte*.

REFERENCES

[1] L. B. Alberti, *De re aedificatoria*, libro IX, capitolo V, Milano: Il Polifilo, 1966

[2] M. Michelutti, *Tra musica e architettura*, Conservatorio di Musica di Milano, 2003, (http://digilander.libero.it/initlabor/musica-architettura-michelutti/musica-architett-marta1.html)

[3] S. Hall, "Stretto house", in *Pamphlet Architecture 16 - Architecture as a Translation of Music*, E. Martin, Ed. New York: Princeton Architectural Press, 1994, pp.56-59.

[4] K. Mavash, "Site + Sound : Space" in *Resonance essays on the intersection of Music and Architecture,* vol. 1, M. W. Muecke, M. S. Zach, Eds. Ames, IA: Culicidae Architectural Press, 2007, pp 53 -75.

[5] A. Di Scipio, *Using PD for Live Interactions in Sound. An Exploratory Approach*, 2006, http://lac.zkm.de/2006/ presentations/ lac2006_agostino_di_scipio.pdf

[6] A. Di Scipio, "Iterated Nonlinear Functions as a Sound-Generating Engine" in *Leonardo*, Vol. 34, No. 3, R.F. Malina, Ed. Cambridge, MA: Mit Press, 2001, pp. 249–254.

[7] A. Taroppi, *Composizione Musicale con gli Algoritmi Genetici – GA: una libreria Open Music per la composizione musicale assistita basata su tecniche derivate dagli algoritmi genetici*, thesis "triennio di Musica Elettronica e Tecnologie del Suono", Como, IT: Conservatorio di Musica, 2005, unpublished.

[8] P. Ferretti, *Estetica Gregoriana*, Roma: Pont. Ist. Di Mus. S., 1934.

Spatial Orchestration

Eric Lyon
Sonic Arts Research Centre
School of Music and Sonic Arts
Queen's University Belfast
Belfast, UK
e.lyon@qub.ac.uk

Abstract — **The emergence of multiple sites for the performance of multi-channel spatial music motivates a consideration of strategies for creating spatial music, and for making necessary adjustments to existing spatial works for performances in spaces with significantly different acoustic properties and speaker placement. Spatial orchestration is proposed as a conceptual framework for addressing these issues.**

I.Introduction

Projection of electronic sound over multiple speakers has been an aspect of both live and fixed media electroacoustic music since the emergence of these artistic practices. The Telharmonium, perhaps the earliest electronic performance instrument, first publicly presented in 1906, produced its sound through acoustic horns distributed throughout the performance space [1]. The premieres of early electroacoustic masterpieces such as Karlheinz Stockhausen's *Gesang der Jünglinge* in 1956 [2] and Edgard Varese's *Poème Électronique* in 1958 [3] both surrounded the audience with speakers. Spatial projection is the central feature in performances at Audium, which have been ongoing since 1960 [4]. In the case of both *Poème Électronique* and Audium, the performances were site specific, so that the problem of transferring the piece to different spaces did not yet arise.

II.Mapping Strategies

Mapping of channels has been treated flexibly in both directions. For distribution purposes, both the five-channel tape of *Gesang der Jünglinge* and the four-channel tape of *Poème Électronique* were mixed down to stereo for commercial distribution on vinyl recordings. The four channels of *Poème Électronique* were performed into a 400 speaker space, the Philips Pavilion. The source tape for the 136 speaker space of Audium is also a four-channel recording. While mapping from multi-channel down to stereo is inevitably seen as a compromise, mapping from fewer to more speakers is an opportunity. We first consider existing mapping strategies before moving on to spatial orchestration schemes.

III.Diffusion

Diffusion is a key performance strategy in which relatively few channels of audio are mapped to a potentially large number of speakers in a given space. This has the advantage of separating the structure of the original audio from that of the performance space. Diffusion, properly done, can add a sense of liveliness to the composition. Irrespective of locative aspects of spatialization, a good diffusion performance can create the impression that the musical source is comprised of many more tracks than the stereo pair that often forms the basis for diffusion. This is often enhanced by interactions between the recorded audio materials and frequency-specific radiation properties of the performance space.

IV.Ambisonics

Ambisonics encoding attempts to provide compositional access to spatial imaging throughout a given performance space [5]. Perception of localized images should be irrespective of listener position. However, as part of the ambisonics process, input sounds are filtered, which may go against the composer's intentions. Speakers must be placed fairly precisely in the space in order to preserve the ambisonics effect.

V.point Source

Point source composition requires advance knowledge of the locations of speakers in the space. The spatial location of each sound is then calculated as part of the composition process. This is a highly effective method for realizing trajectory patterns. Point source is essentially a panning scheme, without the psychoacoustic filtering in ambisonics.

VI.Current standards

It is customary at electroacoustic music festivals to have access to at least an eight-channel (octophonic) playback system of good quality. However there are two common octophonic configurations, box and diamond, and neither configuration appears to be winning out over the other. One could compose with either point source or ambisonics for up to eight speakers, and expect a reasonable performance, though the ambisonics encoding would still be subject to vagaries of speaker placement to a greater degree than point source. Diffusion can be done to any number of speakers. However the more speakers available, the more difficult to control the performance with live diffusion, though this problem can be addressed with multiple performers, if a software system is prepared for merging control streams to the computer controlling outputs. But in general diffusion will still be limited to spatial movements that can be performed in realtime. Spatial aspects involving fragmentation of the sound, for example spectral diffusion could not be performed with precision, without custom software.

VII.SPATIAL ORCHESTRATION

We do not have a good model for composing for performance of music with greater than eight channels into a space other than that for which the piece was composed. Spatial orchestration is proposed as a set of strategies to meet the challenge of composing for larger numbers of speakers. It is not a magic bullet that automates a mapping process. Instead it suggests that composers re-orchestrate spatial attributes of their piece for each new space in which performances will take place. Rather than compose relatively few tracks to be expanded for each multi-channel space by means of live performance, the composition should make full use of the spatial properties of the hall for which it is composed, most likely creating separate tracks for each available speaker.

VIII.BASIC MAPPING

The Sonic Lab at the Sonic Arts Research Centre can project sound from at least three elevation levels: below the audience, at audience level and above the audience. Works composed to be performed in the Sonic Lab can take advantage of these properties and design pan patterns that move above and below the audience, as well as around them. However, most other multi-channel spaces do not have the capability to project sound from below the audience. Thus a work composed for the Sonic Lab would need to have its spatial properties adjusted when performed into another space. The elevation panning might be preserved, but it might start at ground level and then pan above the audience. Similarly, if a work composed at Sonic Lab is projected into a space with two layers of elevation, the three layers of Sonic Lab elevation could be recompiled with a virtual layer halfway between the upper and lower elevation level.

IX.OTHER ASPECTS TO REORCHESTRATE

While all the speakers in the Sonic Lab are placed around the audience, in BEAST performances and MANTIS performances, speakers are placed throughout the space. In this case, surround trajectories may be reconceived to move throughout the space. Or existing virtual motion through the space could be re-orchestrated into actual point source motion through speakers distributed throughout the audience. Another important aspect of the space will be its reverberant properties. In case of moving a composition from a fairly wet space to a drier one, convolution with impulse responses from the original space might preserve the intended spatial environment. Alternatively, the actual timing of events might be changed as appropriate for the new performance space.

X.PRACTICAL ASPECTS OF SPATIAL ORCHESTRATION

Since the strategies of spatial orchestration require a degree of experimentation in each target hall, it is proposed that a group of multi-channel works be composed at a consortium of institutions with multi-channel performance spaces. Each work should be performed in each space, with a suitable amount of time given to prepare the piece for the space. This would require not just sufficient rehearsal time, but a sharing of essential aspects of the hall in advance of the performance, possibly by a preliminary composer residency, so that the work can be properly prepared for each performance.

XI.COMPOSITIONAL STRATEGIES

Compositional strategies are highly conditioned by the outlook of the individual composer, as well as the aesthetic goals of specific compositions. The constraints of the compositional environment further condition these strategies. Composing multichannel works with precise intended spatial attributes imposes serious constraints, often in the form of stringent requirements for the performance situation, especially when ambisonics encoding is involved. Composing multichannel works that are intended for multiple performances imposes the potential constraint that the piece will be restricted exclusively to the compositional possibilities that are most likely to robustly survive significantly different spatial performance environments. Spatial orchestration eases this last constraint, with the understanding that the composition may be reconfigured for each individual performance. This will require at times that the composer conceive of spatial attributes in a more abstract fashion, that is then instantiated in potentially different ways into different performance environments. The key practical measure will be to maintain the composition in highly multichannel fashion, such that the individual elements can be remixed, or reconstituted for each performance as necessary.

XII.FUTURE WORK

The work presented here is ongoing, as it is in the early stage of conception and experimentation. At present all experimentation has taken place at the Sonic Arts Research Centre. The next step will be to compose a complete multi-channel work, intended for performance in multiple venues. Several collaborating institutions have been identified, and it is anticipated that this work will commence in 2009.

REFERENCES

[1] Crab, S. 2005. *Thadeus Cahill's "Dynamophone/Telharmonium"* http://www.obsolete.com/120_year/machines/telharmonium/ (13 April 2008).

[2] Smalley, J. 2000. Gesang der Jünglinge: History and Analysis. http://www.music.columbia.edu/masterpieces/notes/stockhausen/GesangHistoryandAnalysis.pdf (15 April 2008).

[3] Mondloch, K. "A Symphony of Sensations in the Spectator: Le Corbusier's Poème électronique and the Historicization of New Media Arts", *Leonardo*, **37**(1), 57-61, 2004.

[4] Loy, G. "About AUDIUM, A Conversation with Stanley Shaff", *Computer Music Journal*, **9**(2) 41-48, 1985.

[5] http://www.ambisonic.net (13 April 2008).

Applications of Typomorphology in *Acute*; Scoring the Ideal and its Mirror.

Ricardo Climent
Novars Research Centre, University of Manchester, Coupland Street, M13 9PL
Manchester, UK.   ricardo.climent@manchester.ac.uk

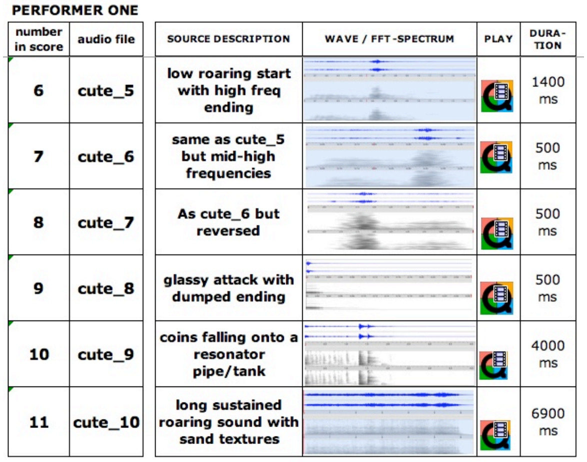
***Abstract* — Acute, is a music score composed by the author of this paper, for Percussion Quartet and Fixed Media (tape) using 'Searched Objects' as instruments. This paper examines how this piece recontextualises existing research in Typology and Morphology of Sound Objects to produce a unique music mixed-media score, for the exploration of the sonic possibilities when confronting the 'ideal' (sonic object to be found) with 'the reconstruction of itself' through time (when performers attempt to recreate the given sounds) using processes of Spectro-gestural mimesis.**

I. Reconstruction of Instrumental Forces

A. Found Objects vs. Searched Objects:

The concept of 'Found Objects' has been extensively explored and discussed in composition and improvisation environments, especially in percussion pieces [1]. However, by introducing the idea of 'Searched Objects' as an extension of it, I was confident that there was scope for another twist, if methodologies and creative thinking from the field of Acousmatic Music and experimentation with performers' sonic memory and experience were introduced. To explain what Searched Object means in Acute's context, we start with the music score, the notation of which is totally fixed but the instrumentation not physically determined. In other words, the composer has not specified what exact percussion resources have to be used. Instead, the instrumentation is sonically provided on a Compact Disc and a chart, as a collection of 44 very short sound samples recorded and edited on a computer. Those sounds are strategically distributed among the four percussionists.

TABLE I.
Excerpt of a Chart Provided to Performer One, Including Embedded Audio, Sonogram and Description of the File



PERFORMER ONE

number in score	audio file	SOURCE DESCRIPTION	WAVE / FFT-SPECTRUM	PLAY	DURA-TION
6	cute_5	low roaring start with high freq ending			1400 ms
7	cute_6	same as cute_5 but mid-high frequencies			500 ms
8	cute_7	As cute_6 but reversed			500 ms
9	cute_8	glassy attack with dumped ending			500 ms
10	cute_9	coins falling onto a resonator pipe/tank			4000 ms
11	cute_10	long sustained roaring sound with sand textures			6900 ms

Kontakte Percussion Group [2], the performers who commissioned this electro-acoustic score, had therefore to make extensive use of their acute sense of hearing and their prior performance experience by calling on their sound memory and creativity. After listening to the sample, they had to 'Go and Search' for the necessary physical objects and 'Invent' the gestures and microphone techniques to reconstruct each of the forty-four samples, in order to determine the instrumental forces, (thus the term Searched Objects was created).

During this process of reconstruction, which involved microphone experimentation and techniques, performers were assisted by a sound engineer, who also produced the CD recording of the piece in the summer of 2007.

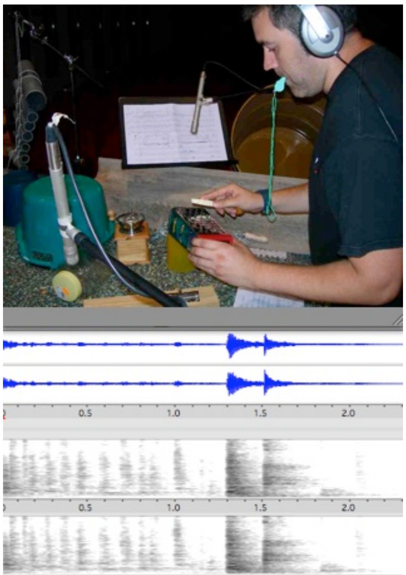


Fig. 1. A more detailed sonogram and ensemble leader (Miguel Angel Orero) reconstructing the gesture in the sonogram using searched objects

II. Applications of Typomorphology; a Step Further

A. Score Recontextualisation

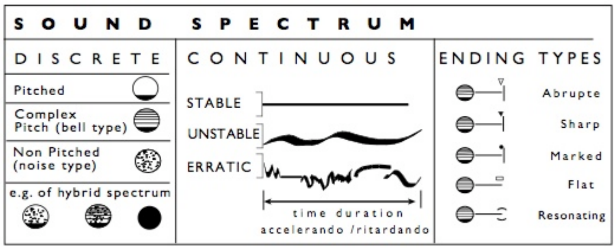
This score is a recontextualisation of existing research in Typology and Morphology of Sonic Objects started by Pierre Schaeffer in Traite des objets musicaux, 1966 [3], continued among others by Dennis Smalley, Spectromorphology, 1997 [4], and reunited by Lasse Thoresen, Spectro-morphological Analysis of Sound Objects, 2001-04 [5].

The musical score, including the computer part (fixed media) is a journey for creative exploration of theoretical concepts such as Sound Spectrum, Spectral Brightness and Pulse, using live sound objects which function as another layer of a fixed media part. The fixed element (the tape) was created exclusively with the already mentioned 44 very-short materials, aiming to find beauty and musical expression in the matches and divergences between the original recording (the ideal) and the mimetic sound designed live by performers (the imitation of the ideal).

B. *From Spectromorphological Analysis to Notated Sound*

Although Thoresen developed this theory of Spectromorphological Analysis of Sound Objects for the analysis of acousmatic music pieces, my intention was to reorient his findings towards writing a musical score (notated sound). I wanted to describe in an precise way, how after solving the Searched Objects puzzle, the final instrumentation should be intervened and how performance would highlight timbral, textural and morphological aspects of the sounds, which finally constituted the heart of the music composition.

TABLE II.
CLASSIFICATION OF SOUND SPECTRUM NOTATION IN *ACUTE*



SOUND SPECTRUM		
DISCRETE	CONTINUOUS	ENDING TYPES
Pitched	STABLE	Abrupte
Complex Pitch (bell type)	UNSTABLE	Sharp
Non Pitched (noise type)	ERRATIC	Marked
e.g. of hybrid spectrum	time duration accelerando /ritardando	Flat
		Resonating

Table 2 Score legend is a chart for performance instructions to define Sound Spectrum, based upon typomorphology vocabulary.

C. *Scoring the 'ideal' and the 'copy of the ideal' thoughout the Typomorphology of Sound Objects*

The idea of exposing the 'ideal' or 'model' (the recorded sound provided) to its closest possible 'copy' (the encountered object and methodology to reproduce the original sound) is informed by early pioneer Film work by Andrej Tarkowskij [6][7], (Solaris, der Spiegel etc). As a matter of fact, what Tarkowskij does is to contextualize in his media the theories of Marx and Hegel, where Marx criticises Hegel's thinking that the 'ideal' determines the 'material', and suggests the inversion of the primacy of the 'ideal' (consciousness, thought, ideas), over the 'material' (world). In other words, the material determines the ideal and not vice versa (opposite to the starting point of Acute). Marx goes beyond his own thinking, and denies the very existence of the 'ideal' as a separable entity [8].

In Acute, the 'ideal' is given, (the 44 brief samples' guide to reconstruct the instrumentation) and it must determine the 'material' (physical objects / sound sources allocated to attempt to reproduce the 'ideal') as in Hegel.

To give another twist to the theory, I consciously transformed some of the given brief samples with loose reference to its source and made them nearly impossible to reconstruct by acoustic means. As with Marx's ideas, I wanted to question the possibility of achieving the 'ideal' by creating it as some sort of abstract form but still scoring it as precisely as possible, using the principles of typomorphology.

The score pushes this strategy even further and makes its sonic discourse out of *sculpting in time* [9] both 'source' and 'mirror' and exploring the thin line between the two.

In this electro-acoustic music composition I wanted to investigate the scope for sound exploration when chasing the creative sonic possibilities resulting from a process of spectro-gestural mimesis between the computer-generated sample and the acoustic sound imitating it. I found refinement of musical expression in the divergences between 'model' and 'copy' but also in the similarities and the coherence between the acoustic and fixed media materials (the tape).

D. *Notated methodology; a brief explanation:*

Performers need a headphone monitoring system to hear the click track and voice track. In Figure 2 below, Mk_i cues are rehearsal marks and key points of synchronization between tape and instruments. P1 (Performer one), will play the set of Searched Objects number 8 (which refers to sample 8 reconstructed by the performer) at 1 min. and 10 sec. The little triangle attached means to execute 8 with a *Sharp Onset*. Dynamics are forte to piano subito and a pause is needed. The sound spectrum to be achieved should have more noise content than pitch. Then object 8 needs to be performed in ascending tremolo and dynamics for less than two seconds. After 1:13, the same kit (8) should be performed but searching for a sound richer in pitch and harmonics (the kit should have both choices) and should be executed creating an 'unstable' continuo for 4.5 seconds.

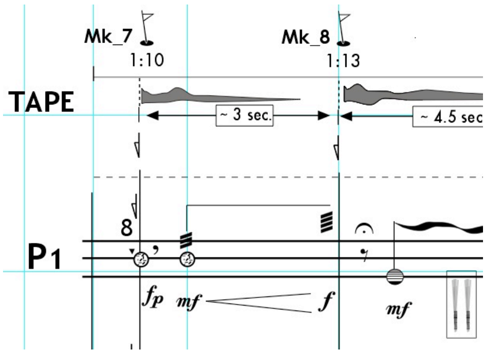


Fig. 2. Score except including the Tape part and one instrument

E. *Structural Rhythm*

The articulation and structure is entirely driven by rhythm. However, it is not just understood as rhythm in musical terms but as *pace*, informed by Tarkowskij's methodology, which is not focused on the Temporal Editing but on the Rhythm of the Scenes (sculpting in time); in Acute terms, it refers to the pace of different Sonic Scenes with a characteristic typology.

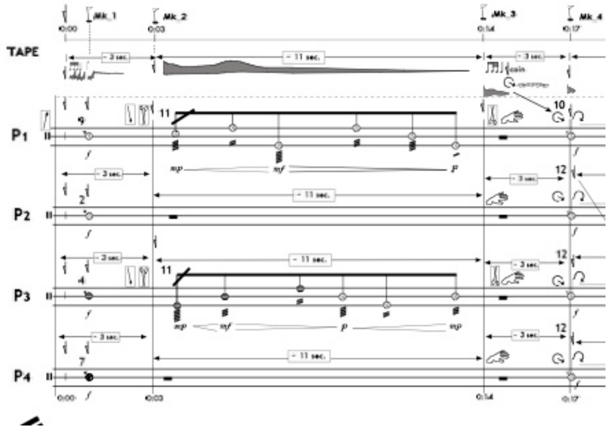


Fig. 3. Score: Sonic Scene excerpt of Acute with an embedded pace.

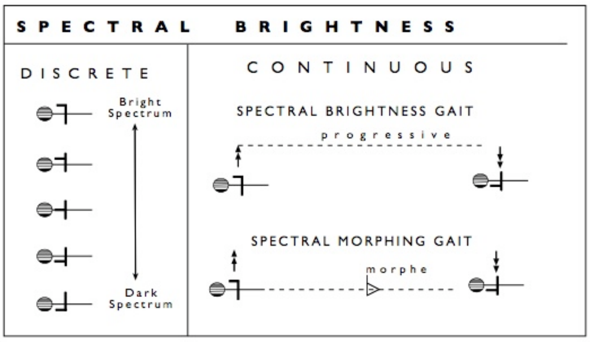
The full score on pdf format and an mp3 version of the piece can be downloaded from here [10].

F. *Aspects of Typomorphology being observed and notated:*

- Sound Spectrum in discrete and continuous form explores different degrees of pitched, non-pitched and complex-bell-like sounds, including onset of offset variations, (see Table II).

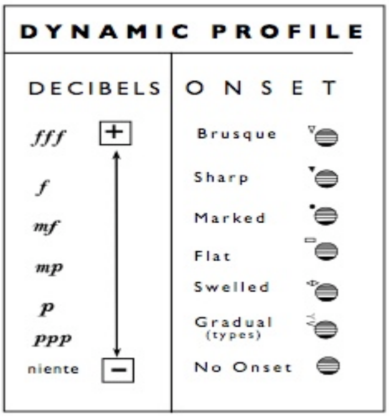
- Spectral Brightness (including degree of darkness and morphing in discrete and continuous form)

TABLE III.
CLASSIFICATION OF SPECTRAL BRIGHTNESS IN *ACUTE*



- Exploration of dynamics (including variations of Onset/ offset) and degree of Textural Granularity, among others.

TABLE IV.
CLASSIFICATION OF DYNAMIC RANGE AND ONSET IN *ACUTE*



- Interlocking and Gesture: between live percussion sounds (imitating the sources) and fixed media (constructed with those sources). Sculpting-time techniques included strategies to blur the differences between the 'image' and its 'mirror' and to explore musical expression in the divergences.

III. CONCLUSIONS

This paper has discussed the implementation aspects of Schaefferian Typomorphology in a mixed-media piece for percussion called Acute. With the introduction of the concept of Search Objects (after Found Objects) in this context, the author proposes a new route for creative expression, informed by Tarkowskij's ideas and methodologies in Film, when exposing the *Ideal* to *its Mirror*. Similarly, the score utilizes vocabulary and grammar from areas of Spectromorphology to precisely notate sound and gestures, which imitate the original aural models generated and transformed employing computer and recording techniques. An explanation of the existing methodology to read and interpret the score leads to some detailed examples about how to notate in time concepts such as Sound Spectrum, Spectral Brightness and others which belong to the Schaefferian vocabulary.

ACKNOWLEDGMENT

Acute was commissioned by Kontakte Grup the Percussió and was released on CD by the IVM label, Valencian Government, Spain in May, 18th, 2008 (Deposito Legal V-1781-2008. The score and research behind it was realised between 2006 and 2007.

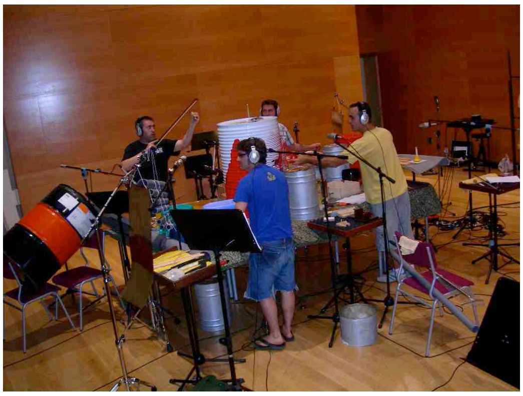


Fig. 4. Recording Session of Acute. Kontakte group reconstructing Sound Sources as Instrumental Forces

REFERENCES

[1] Percussion in the New-Old World. Wilfrid Mellers. The Musical Times, Vol. 133, No. 1795 (Sep., 1992), pp. 445-447.

[2] Kontakte Grup de Percussió. http://www.kontakte-percusion.com. Last visited 28-April-2008.

[3] Typology and Morphology of Sonic Objects. Pierre Schaeffer . Traite des objets musicaux,. Essay Interdisciplines. Nouvelle edition, Seuil. 1966

[4] Dennis Smalley, Spectromorphology: Explaining sound-shapes. Organised Sound: Vol. 2, no. 2. Cambridge University Press: 107-126. 1997

[5] Spectromorphological analysis of sound objects: an adaptation of Pierre Schaeffer's typomorphology. Lasse Thoresen. Organised Sound archive. Volume 12 , Issue 2 (August 2007). Pages 129-141. 2007

[6] Andrej Tarkowskij's films. Lost Harmony: Tarkovsky's "The Mirror" and "The Stalker" Michael Dempsey. Film Quarterly, Vol. 35, No. 1 (Autumn, 1981), pp. 12-17. Published by: University of California Press

[7] Back to the House II: On the Chronotopic and Ideological Reinterpretation of Lem's Solaris in Tarkovsky's Film. Roumiana Deltcheva and Eduard Vlasov. Russian Review, Vol. 56, n.4 (Oct., 1997), pp. 532-549. Published by: Blackwell Publishing on behalf of The Editors and Board of Trustees of the Russian Review

[8] Marx's Idealist Critique of Hegel's Theory of Society and Politics. David A. Duquette. The Review of Politics, Vol. 51, No. 2 (Spring, 1989), pp. 218-240. Published by: Cambridge University Press for the University of Notre Dame du lac on behalf of Review of Politics

[9] Sculpting in Time: Reflections on the Cinema. by Andrey Tarkovsky; Kitty Hunter-Blair. Slavic Review, Vol. 48, No. 2 (Summer, 1989), pp. 348-349. Published by: The American Association for the Advancement of Slavic Studies

[10] http://spatialisation.7host.com/acute/acute.html [18/06/08]

Space Resonating Through Sound

Fernando Iazzetta*, Lílian Campesato†

* University of São Paulo/ Music Department, São Paulo, Brazil, iazzetta@usp.br

† University of São Paulo/ Music Department, São Paulo, Brazil, lilicampesato@gmail.com

Abstract — **In this paper we will analyze how the conception of space in music is expanded by the repertoire of sound art, moving from the idea of space as a delimited area with physical and acoustical characteristics, to the notion of site in which representational aspects of a place become expressive elements of a work of art.**

I.INTRODUCTION

During the last century, sound has followed two different paths within the arts. On one hand we have the musical route toward the creation of strategies and procedures to establish a self-referential grammar. This process culminates with the emphasis on the sound itself as a source and foundation of musical discourse as one can find in Pierre Schaeffer's *écoute réduite* [1] or in John Cage's transpositions of everyday sounds into music [2]. On the other hand we saw, especially outside the musical domain, a tendency to explore the contextual and representational potentiality of sound that is in the basis of an expressive part of what constitutes the repertoire of today's sound art. In the first case comes out the idea of musicalization of sounds [3]; in the second, it becomes manifest a direct connection of sound material with other aspects of culture and life.

As a consequence of this second instance, many artistic manifestations in which sound played a significant role have to shape new relations between sound and the referential world where sound takes place. This is the case, for example, of installation art, performance art and sound art. One of the references that emerge in this context is a notion of place that transcends the idea of geometric space as a measure delimited by geographical coordinates [4]. It incorporates perceptual, social, psychological, acoustical and visual characteristics of an ambient assigned by specific circumstances and occurrences.

Thus, site-specific becomes the conception shaping the work in which the notion of space embraces more than geometrical properties: materials as well as the history they can elucidate, architectural contexts, and even the cultural and social conventions that regulate the place of exhibition, they all became constitutive elements of the art work.

As Guy Lelong [5] points out, every art form needs a place to happen – the book, the gallery, the museum, the concert hall – and the relationship between the work of art and it's place can lead to two types of reflection. In the first case, the place is the device intended to present the work. It can be understood as a frame that both draws up the boundaries of the work and restricts its existence. In the second case, there is the idea of art *in situ*, in which the site transcends the perspective of a frame that traps the work to become part of the work.

When the site becomes part of the artwork it is converted into a space that is more than the place where that artistic object is presented. It brings different aspects of the environment such as architectural features, social conventions, informational traces and curatorial characteristics to the compositional level of the work. This practice introduced a criticism regarding the modes of art diffusion and lead to a shift from a focus aimed at the object to a wider concern with the environment.

In the case of music, the concert hall is such a consolidate space that it became extremely difficult to project new musical environments and new forms of music presentation. Even some more sophisticated proposals of electroacoustic multitrack diffusion[1] did not break up with the formal configuration of a concert hall in which space serves as an enclosure where the music physically is constrained and the audience is confined.

When music crosses the boundaries of the concert hall in search of alternative spaces it is usually reconfigured in new formats such as sound art, sound installations and performances. The art gallery became an alternative space to host musical and sonic arts. But if the concert hall enforced its own ritual and traditions, the gallery also provided a new type of space with its own conventions. As in the concert hall, the gallery walls and rooms also impose a clear delimitation of what is inside and what is outside.

II.SPACE IN MUSIC AND SOUND ART

The art gallery has established itself as an almost neutral and aseptic space, but sealed (without windows), especially built to isolate the work of art from any external event: a white cube.

> The ideal gallery subtracts from the work of art all the evidences that interfere in the fact that it is 'art'. The work is isolated from anything that might undermine its appreciation. This provides the room with a characteristic presence of other spaces in which conventions are preserved by the repetition of a closed system of values. Some of the holiness of the church, of the formality of the court, of the mystique of the experimental laboratory is joined to a fashionable design to produce a unique chamber of aesthetics [6, p: 3].

As formal institutions the white cube and other sites like the concert hall, clearly establish a separation between

[1] For example, in 1958 Edgard Varèse's *Poème Èlectronique* was diffused over more than 400 loudspeakers at the Philips Pavillion designed by Xenakis and Le Corbusier for the Brussels World's Fair; also, in 1970, at the Osaka Expo, Karlheinz Stockhausen performed electronic compositions at the German Pavillion, a spherical auditorium, equipped with over 50 loudspeakers in concentric rings around the audience.

the viewer / listener and the work, as well as establish the annulment of the body and of the presence of people:

> Certainly the presence of that strange piece of furniture, your own body, seems to be superfluous, an intrusion. The site raises the idea that while eyes and minds are welcome, the bodies that occupy space are not - or are tolerated only as synesthesic dummies for future study [6, p: 4].

It is interesting to notice that in the case of installations, including sound installations, the institutions constituted around the concept of white cube (galleries, museums) are changed into a kind of black cube, a site composed by dark surfaces (walls). At the same time that it is set as an isolated environment the black cube eliminates the delimitations of space, transcending the sacredness of the white cube, and involving the viewer in an immersive manner [7]. Black cubes are inhabited by loudspeakers and screens, and the dark walls become invisible, providing the creation of a virtual space in which new perceptive modalities are stimulated.

The gallery's white walls lead to a contemplative attitude from the spectator, who, framed by social rules, keeps a relative distance from the work, establishing a relationship that is more rational than physical, corporal or sensorial [6, 7]. The contemplative detachment is clearly a remnant of an almost religious attitude, in which the aseptic white cube and the concert hall are part of an almost sacred conception of art.

The installation attempts to transform the "white cube" into a "back cube": as it weakens the idea of sacred, it brings the viewer to a closer relation with the work, and transforms the site into an involving environment [7]. Thus, the context becomes content and the viewer becomes part of the work, emphasizing the idea of immersion. Material elements and creative procedures are transformed.

In the installation, both music and sound art are released from the time of performance, even from the time of recorded performance. The time of the work becomes the time the spectator, who may decide when to start and stop paying attention to every aspect of work. This penetration in the space of the installation consists in a significant shift in the modes of music production.

The installation is configured as a new way of presentation of the work of art. The listener is not confined to a fixed position in space, but is invited to create his/her own spatial relationship with the work [7]. While the way of publishing music is traditionally achieved by performing it (or at least by recording its performance), installation art configures itself as "the possibility of publishing [the music] without performance" [8] in a way that the listener is free to establish his/her own relations with the time of the work and with the space where it happens.

One of the transformations operated by electroacoustic music in the mid-twentieth century was the appropriation of space as a musical element. However, in this process there was a reduction of the idea of spatialization to the concept of a projection of sounds in space. Thus, the electroacoustic project devoted a considerable attention to the development of strategies to create the perception of sound source localization (front/back, left/right) and of sonic planes (close/distant), as well as to produce virtual acoustic spaces (small/large, dry/reverberant). In this sense, the acoustic space in electroacoustic music is mainly related to sound data and the aural perception of space lies upon source localization and room dimensions. It constitutes an *acousmatic* space [9] that does not correspond to the space where the music is diffused. This virtual space is constructed by the electronic devices ridden behind the curtain of loudspeakers.

In sound art, the real space where the work takes place is part of the work. Multimodal sensations are activated by acoustical elements and space resonates with sound, leading to sensorial images of dimension, color, texture, shape, and movement. Also, many of the references attached to a site can be triggered by sounds and become part of the work. The idea of space is translated into the idea of site, incorporating social, psychological, perceptive, acoustic and visual characteristics of a place. Space becomes a representational element in the artwork.

III.THREE INSTANCES OF SPACE IN MUSIC AND SOUND ART

Any art form that takes sound as it main material is a temporal art form since sound can only happens in time. In music and in other types of sound art, time can be associated with space forms in different levels of relevance. We can analyze some of the spatial aspects that can be put in resonance by sound in music and sound art, showing how they are explored in recent works. For this purpose we will establish three categories of spatial relationship with sound and analyze their role in the art. The three categories are: acoustical space; architectural space; and representational space.

A. Acoustical Space

Acoustical space embraces the perceptual acoustic characteristics of space, such as volume, reverberation and sound source localization. Acoustic phenomena are used to provide psychoacoustic impressions of a space. This instance is related to what is generally called spatialization in electroacoustic music or stereophonic image in the process of sound recording and mixing in a studio. Basically, it is related to the localization of sound sources and their movements in space. Also, the reverberant characteristics of a perceived sound allows the perception of some spatial aspects such as volume, shape and can even provide some hints about the type and position of surfaces constituting that space.

Although we can cite some examples of use of spatialization in different periods of music history[2], space only becomes to be considered as a compositional element on the second half of twenty century. Electroacoustic music has included the use of spatial diffusion since its very beginning. Multi-channel recording and the positioning of multiple loudspeakers around the audience became an essential part of electroacoustic compositional agenda since the creation of the *Gesang der Jünglinge* (1955-56) by Karlheinz Stockhausen. As the composer remarks, "the function of space has been neutralized in our western music" [10, p: 101], but with electronic music the displacement of sound in space and the projection of sound in different planes of distance became as important as other compositional elements: "Building spatial depth

[2] For example, the contrapunctual distribution of choirs in the San Marcos' Cathedral by Giovanni Gabrielli (1557-1612); the placement of musicians behind the stage in Gustav Mahler's *Second Symphony* (1894); and the distribution of musicians around the concert hall in Charles Ives' *Unanswered Question* (1908).

by superimposition of layers enables us to compose perspectives in sound from close up to far away, analogous to the way we compose layers of melody and harmony in the two-dimensional plane of traditional music" [10, p: 106].

Following this perspective, in the last decades electroacoustic music has developed many systems and strategies to deal with sound spatialization. From the first experiences with multi-channel composition in electronic music in the 1950's[3] to the set up of large loudspeakers orchestras[4] to the development of new spatialization techniques and protocols[5], space got into the musical agenda.

In sound art, space has become the material that constitutes the essence of many works as they emphasize acoustic effects produced by the controlled projection of sound sources. Not only sound itself is perceived in relation to space, but also the acoustical and psychoacoustical characteristics of sound spatial dimensions are employed to emphasize the architectural characteristics of a particular place. In fact, many sound art works explore psychoacoustical aspects by focusing in the perceptual subtleties of sound events. They seek to stimulate the spectator to understand acoustical phenomena that usually are not taken into account, even in the processes of listening to music [7].

For example, in the installation *Fünf Felder* (*Five Fields*, 2002), Christina Kubisch takes advantage of the room's architecture to organize the space in relation to sound. The room is divided into five fields defined by the window niches. Loudspeakers of different sizes are placed within these fields and diffuse sounds generated by 15 tuning forks covering a spectrum from 64 to 4096Hz. Loudspeakers are painted with a varnish that shines under ultraviolet light. The artist employs the reflection of light to put visible and audible structures in evidence. The perception of radiated lights oscillates between illumination and luminosity, highlighting areas and lines that float around the space of the room, enhancing the amalgam between audible and visual perception [11].

Plight (Anthony d' Offray Gallery, 1985) is an installation created by German artist Joseph Beuys, well known by his ritualistic performances and by his participation in the Fluxus Group. In this installation one can notice the conceptual use of the acoustic space as it uses a highly sound-absorbent material to cover all the surfaces of a room. This material eliminates natural room's reverberation and creates an extreme perceptive distinction between the external and internal acoustic space. Inside the room the sensation is as if all the sounds in the environment have been dragged by the covered surfaces. There is no need of playing specific sounds or music to perceive the acoustical changes: environmental noises or sounds produced by the visitor are sufficient to trigger the attention to the unusual acoustics. Inside the room a piano remains silent as if its sounds were also drawn by the absorbent surfaces (Fig. 1).

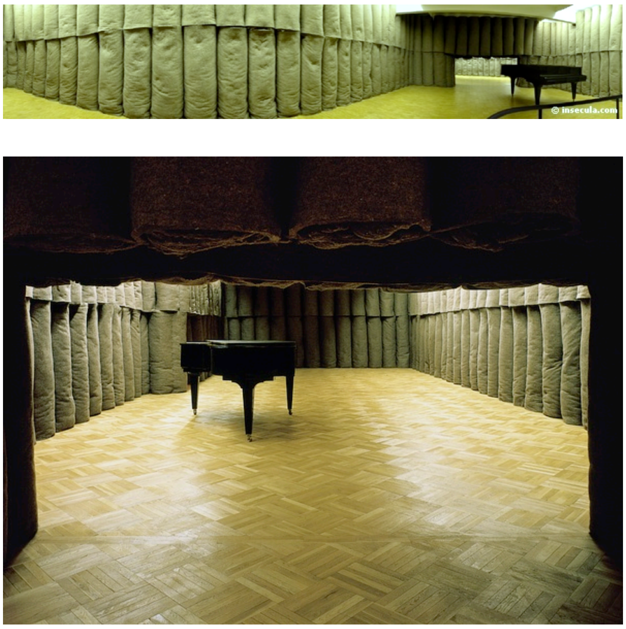


Fig 1: Details of *Pligth* (1985), by Joseph Beuys

In *Stationen* (1992), Robin Minard creates an installation where "space itself becomes a musical instrument and architecture an acoustic event" [12, p: 99]. In this work, loudspeakers are placed around the stairwell and the bell tower of Berlin's Parochial Church (Fig. 2). These loudspeakers reproduce natural and synthetic sounds that are integrated to the acoustics of the environment without disturbing it. Some of the sounds were produced and controlled by a computer and were reproduced by loudspeakers placed in positions of the building following a vertical organization in which the register of the sounds change gradually from low to high as on ascends the stairwell into the bell tower room. The filtering of higher frequencies and the integration of loudspeakers with the environment make it difficult to localize the sound sources providing a very diffuse sound reproduction. As the artist points out,

> The overall dynamic range of the installation was adjusted to always retain the effect of a light, homogeneous spatial coloring which very gradually changed in register as one ascended the stairwell into the bell tower room. The dynamic levels and the harmonic content of separate components of this sound color also varied in real-time, with the aid of a computer, in relationship to the amplitudes and frequencies present in outdoor sounds […] Through such accentuations of the space's acoustic properties, combined with the spatial rather than temporal organization of sounds, the installation created an environment in which sound often drew the listener's attention to *architectural* aspects of the space rather than solely to a specific musical content [12, p: 98-99].

[3] For example, the *pupitre d'espace*, built in 1951 by Jacques Poullin at the studio of the Radiodiffusion-Télévision Française (RTF) for real-time quadraphonic spatialization; and the 400-loudspeaker system set in the Phillips Pavilhon at the Brussels World's Fair for the presentation of Edgard Varèse's *Poème Electronique* in 1958.

[4] Such as the *Acousmonium* developed at GRM in the mid 1970's.

[5] Ircam's Spatializateur, Ambisonics system and wave field synthesis are representative examples.

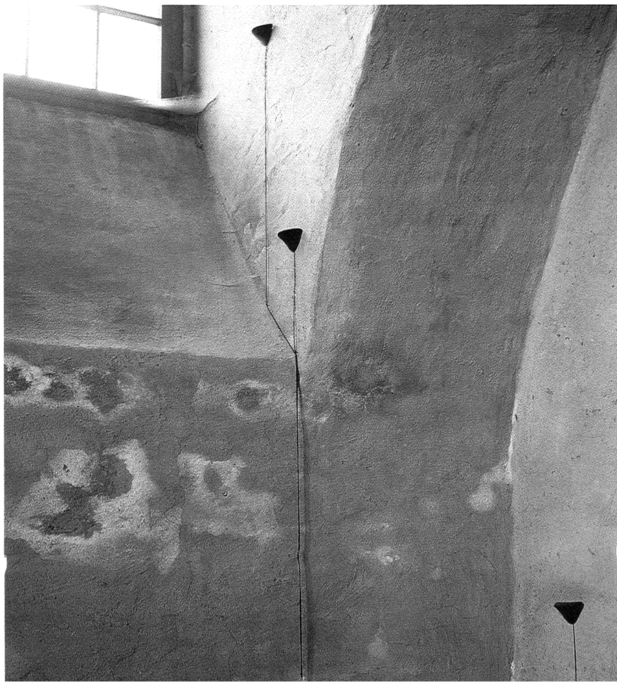


Fig. 2: Detail of *Stationen* (1992), by Robin Minard, Parochialkirche, Berlin.

In both works space is put in evidence by the sounds. On another hand, sound material establishes a relationship of resonance with the space where they are produced. While in music the practice of spatialization remains attached to the idea of sound source localization and displacement, in sound art works space tends to acquire a more effective role by establishing more direct connections between sounds and the acoustic behavior of these sounds in a particular environment.

B. *Architectural Space*

Architectural space relates to the conception of aural architecture developed by Blesser and Salter [13] in which sounds are able to shape a sonic space that carries specific functions and representational potentialities. This conception leads to a close relation between the place where sounds are projected and the way one listens to it, creating a space of listening. In this sense it shares characteristics of both acoustical and representational spaces.

Differently from a soundscape, in which sounds are important in themselves, in aural architecture sounds illuminate space [13, p: 16]. In electroacoustic music the idea of spatialization focus on sounds themselves and on their movement across a virtual space: sounds are put in evidence by their movement. When we think of aural architecture, sonic sources are put to reveal space. In this context space is not taken only as a physical dimension, but also in its social, perceptual and experiential aspects.

According to [13], one can experience space in four modes: "social, as an arena for community cohesion; navigational, as local objects and geometries that combine into a spatial geometry; aesthetic, as an enhanced aesthetic texture; and musical as an artistic extension of instruments" [13, p: 64]. These modes can coexist and their relevance depends upon the cognitive strategies one adopts in a particular context.

Generally, music seldom directs the attention towards space in this sense, even when it incorporates spatial aspects in the compositional process, as we have already mentioned regarding electroacoustic music. Of course, when one listens to a sacred piece inside a church or to an orchestral concert in a park on a Sunday morning, the ambient becomes part of the music and one can establish connections between contextual characteristics of those spaces and the music being performed, but these connections are more accidental than intentional. In this case, the relationship between music and space is more related to the particularity of a performance than to the compositional conception of the work and space is usually drawing attention to musical aspects.

In the field of sound art many works may invert this balance by using sounds and music to emphasize – or to illuminate as would say [13] – the space. Some artists will use sound to put space in resonance and use this resonance to amplify the referential potentiality of space. Robin Minard has created works for public spaces in which the function of music is redefined in relation to noisy environments. He creates a kind of spatial composition for public spaces that receive the artwork without loosing their original functionality. Thus, Minard has "left the protected concert hall to deal with the actual acoustic space of urban world" [12, p: 27], shaping acoustic spaces to become works to be listened to. The use of public spaces requires the creation of strategies to guide the listener's attention to the sounds that compose the natural and urban environment at the same time that the artist "desconstructs and recombines the acoustic material to create an oscillating effect" [12, p: 29] in which one can establish new connections in relation to the familiar sounds that inhabit a place.

In *Brunnen* (1994) on can note the resonance of these ideas. This installation consists of three rectangular blue acrylic boxes asymmetrically located on the floor (Fig. 3). Inside the boxes are loudspeakers that transmit a mixture of natural and synthetic sounds of water. Each of acrylic columns placed close to the speakers are tuned in intervals a quarter-of-tone apart, producing small frequency variations in the environment [12]. It is interesting to note that the work was installed at the entrance of the Mozarteum in Salzburg, whose traditional fountain has been replaced by the blue boxes, producing an interesting integration between the current space and the memory of the previous acoustic space. Moreover, there is a strong spatial relationship created by the formal similarity between the boxes and windows of the building that surround the plant [7].

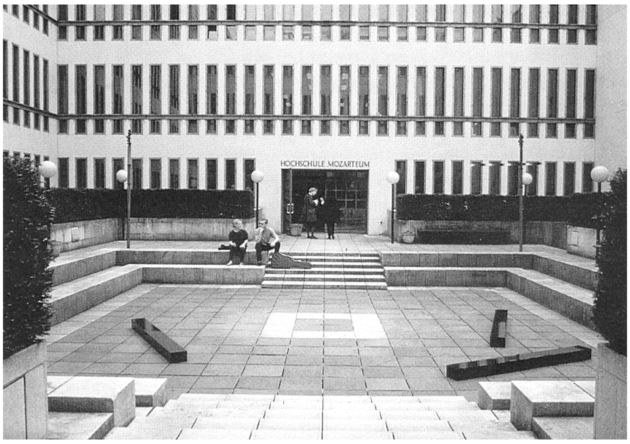


Fig. 3: *Brunnen* (1994), by Robin Minard, Mozarteum, Salzburg.

Generally, in public spaces, the artist must deal with the fact that sounds and other elements of that space carry their own specific meanings, because sounds are attached to the context of their places of origin. In this type of work, Minard adds previously composed sounds to reach a perception that "hovers between identifying familiar phenomena and noting unexpected musical sounds" [12, p: 29].

In *Silent Music* (1994), the artist uses about 400 piezo-electric loudspeakers fixed on walls and other surfaces. Attached to their wires the loudspeakers assume plant-like forms that resemble bioorganic structures [12]. The arrangement of the loudspeakers creates the impression that they search for the light as if they were real plants. The work is conceived for both traditional exhibition spaces and public areas such as gardens and parks. Sounds are composed by synthetic and natural sources and are specially conceived to be incorporated into the environment (Fig. 4).

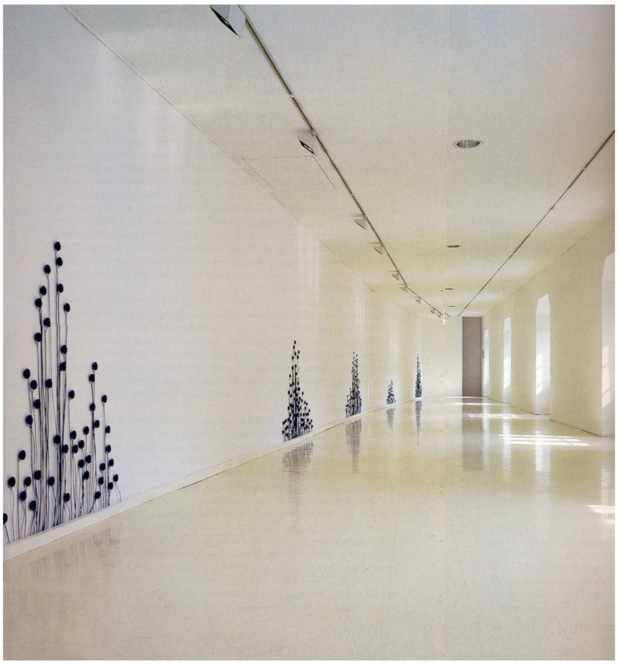


Fig. 4: *Silent Music* (1999), by Robin Minard, Stadtgalerie Saarbrücken.

C. Representational Space

Representational space refers to images, contexts and concepts that are related to a specific site and can be triggered by sounds. It concerns more to the historical and contextual elements of a place than to its the geometrical delimitation and physical configuration. While instrumental and electroacoustic music maintained a discourse based on abstract sound relations that are constructed through references to the musical discourse, sound art tends to generate a representational discourse full of references that point out to concepts and contexts that are external to the work itself. In music referentiality is inserted in the temporal discourse as a basis for the musical narrative. In sound art the referentiality is usually extra-musical, thus it can operate through other resources such as the physical or imaginary space where it is presented. If music tends to establish a linear discourse whose elements are deeply attached to musical grammar, by its turn, sound art uses sound in a more representational way. Thus, its discourse does not need to be strictly based on temporal structures – as is the case of music –, but it can lead to other types of configurations in which conceptual ideas are referred by its sonic constructions.

For this reason time dimension in sound art is somehow condensed. Usually sound art works do not impose a linear temporal organization in which sound elements are strictly distributed in time. Many of these works do not provide a specific begin or end, allowing the spectator to enter and leave the work at any time. As the use of time becomes less imposing, it is possible to adhere to a spatial discourse, typical of the visual arts. In the same way that sound became essential to twenty-century music, space plays a central role in the repertoire of sound art.

An example of a highly representational use of space is the work *Zwölf Türen und zwölf Klänge* (*Twelve Doors and Twelve Sounds*, 2000) by Christina Kubisch. It is part of a series of installations entitled "*consecutio temporum*". Works on this series are created for rooms that went through different historical changes. *Zwölf Türen und zwölf klänge* was set at the second floor of the Opel Villa building in Rüsselsheim [14]. The place was constructed in 1930 and since then it assumed different purposes: it was originally used as a floor for servants, later as hospital and, during World War II, for military purposes (Fig. 5).

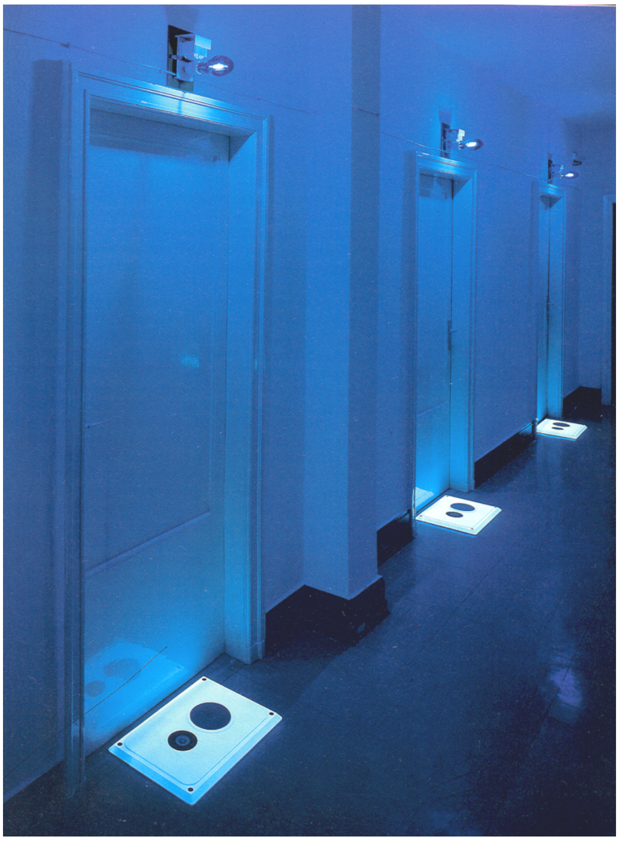


Fig. 5: Detail of *Zwölf Türen und zwölf Klänge* (2000), by Christina Kubisch.

The installation consists of twelve white lacquered doors in front of which are placed twelve white loudspeakers that are made fluorescent by ultraviolet lamps of high intensity. This illumination reveals traces of the building history by making visible some cracks and small marks. These details can guide the public through the past history that emerges from the architecture. It is worth noting the metaphorical use of sounds in relation to the environment. Loudspeakers are placed as thresholds of each lacquered door and each door reflects both the image and the sound of the loudspeaker. The sounds are produced by electronic devices in twelve soundtracks. Their extremely high frequencies are close to the human hearing threshold and work as a tapestry that involves the space.

Regarding the particular Kubisch's conception of an archaeological space in this work, Carsten Ahrens, curator of the exhibition comments that:

> In the luminous dark, the traces of time become visible. Fissures and wounds in the structure of the room's wall appear; our glance and our thoughts follow the patterns of their lines, tracing a journey into what is past. The history of the site becomes a history of question marks, an empty space our curiosity seeks to fill [14, p: 58].

D. Conclusion

In this paper we tried to describe some differences in the conception of space in works of music and sound art. We established three instances in which space is considered in these works: acoustical space, architectural space, and representational space. In these three instances there is a progression from a more objective conception of space to an abstract, referential one. If music repertoire tends to the first instance, sound art is more flexible and explores the potentialities of the three of them in an open manner. As [15, p: 7] points out, "the less musical activity is fixed or centered on the representational handling of representational objects, the more the quest for meaning shifts to the conditions of the social and spatial implications of situations and therefore, of course, space". This also indicates the musical bias toward self-referentiality since music usually employs space to enhance sound qualities. On another hand, sound art exploits referential and representational characteristics of space. Thus, space not only points to internal sonic structures of a piece, but also creates a web of connections among ideas, contexts and stories that lie outside the work.

Acknowledgment

We acknowledge CNPq (Brazilian National Research Council) and Fapesp (São Paulo State Research Fundation) for their financial support.

References

[1] Schaeffer, P. *Traité des Objects Musicaux*. Paris: Éditions du Seuil, 1966.

[2] Dyson, F. "The ear that would hear sounds in themselves: John Cage 1935-1965". *Wireless Imagination: Sound, Radio, and Avant-Garde*. D. Kahn and G. Whitehead. Cambridge, Massachusetts, The MIT Press**:** 373-407, 1992.

[3] Kahn, D. *Noise water meat: a history of sound in the arts*. Cambridge London: The MIT Press, 1999.

[4] LaBelle, B. *Background noise: perspectives on sound art*. New York - London, Continnum, 2006

[5] Lelong, G. "Musique in situ." *Circuit: musiques contemporaines*, vol. **17**(3), pp. 11-20, 2007

[6] O'Doherty, B. *No interior do cubo branco: a ideologia do espaço da arte*. São Paulo: Martins Fontes, 2002.

[7] Campesato, L. *Arte Sonora: uma metamorfose das musas*. University of São Paulo. São Paulo, Brazil, Master Thesis: 173pp, 2007.

[8] Aldrich, N. B. "What is Sound Art?" Retrieved 10/04/2005, from http://emfinstitute.emf.org/articles/aldrich03/aldrich.html, 2003.

[9] Campesato, L. and F. Iazzetta. "Som, espaço e tempo na arte sonora". *Procedings of the XVI Congresso da Associação Nacional de Pesquisa e Pós-Graduação em Música* (ANPPOM), UnB, Brasília, pp. 775-780, 2006.

[10] Stockhausen, K. *Stockhausen on music: Lectures and Interviews compiled by Robin Maconie*. London, New York: Marion Boyars, 1989

[11] Schulz, B., (Ed.) *Resonances: aspects of sound art*. Heidelberg: Kehrer Verlag, 2002.

[12] Schulz, B., Ed. *Robin Minard: Silent Music - Between Sound Art and Acoustic Design*. Heidelberg: Kehrer Verlag, 1999.

[13] Blesser, B. and L.-R. Salter. *Spaces speak, are you listening? - Experiencing aural architecture*. Massachusetts: MIT Press, 2007

[14] Kubisch, C. *KlangRaumLichtZeit*. Heidelberg: Kehrer Verlag, 2000.

[15] Wollscheid, A. "Does the song remain the same?" *Site of Sound: of architecture & the ear*. B. LaBelle. Los Angeles, Errant Bodies Press, pp. 5-10, 1999.

Space as an Evolution Strategy.
Sketch of a Generative Ecosystemic Structure of Sound

Massimo Scamarcio
Scuola di Musica Elettronica, Conservatorio di Napoli, Italy
massimo.scamarcio@gmail.com

***Abstract* — This paper discusses a generative, systemic approach on sound processing, touching topics like genetics, evolutionary programming, eco-systemic interaction of sound in space, and feedback, putting them in the context of the author's *Syntáxis(Acoustic Generative Sound Processing System, part 1)*: a sound installation for stereophonic speaker system and microphone. The main implications of the overall structure of the installation are analysed, focusing on the dynamics of it and its relationships with space. The paper also illustrates the main structure of the algorithm regulating the installation behavior, along with brief references to the software platform used to develop it (Max MSP 5).**

I. Introduction

The concept of space has played a crucial role in the development of art during its history. Implications have been numerous, and while they fall mainly – and understandably – in the realm of visual arts, the full spectrum of their aesthetical, technical, conceptual and creative elements can be found in the framework of music as well. In this sense, the attitudes of visual art and music, regarding space, can be seen as similar in scope: with the dawn of the twentieth century's european avantgarde movements, both aimed at a progressive delinearization of the previous, standardized structures. Breaking down those elements, in music, which were considered obsolete meant, also, reconsidering the space of music itself – which had been up to then historically encased in a left-to-right linearity. The possibilities given by the electro-acoustic research, concrete and synthetic alike, permitted the investigation of space as a compositional element, offering peculiarities and potential to be experimented with, along with those offered by pitch and timbre. The relative implications became too numerous to list here, but they swiftly encompassed a wide range of different forms of electro-acoustic experimentation and research: one of them, the famous Soundscape scene, pioneered by canadian composers R. Murray Schafer and Barry Truax, insisted on a perspective on sound which is coherent with the environment in which is produced. A view which is ecologic in nature, where environment/space is treated as an eco-systemic source of sound, linked with the elements which form and give influence to the system itself, regulating its dynamics. Elements which are the basis of the author's *Syntáxis(Acoustic Generative Sound Processing System, part 1)*, which will be illustrated here along with its design philosophy.

II. Space, Environment, Ecosystem, and their potentials

Any environment in which sound occurs spins a complex web of relationships that are systemic in nature. A sound produced in a determinate environment is inevitably bound to interact with the system and the agents that generated it, carrying along informations about the space in which it occurs: an *acoustic* space, with precise physical characteristics (i.e. specific resonances, attenuations, size, reverberation and so on) which define the sound that propagates, evolving, through it[I]. The relationships between sound in space and the agents which produced it is, therefore, mutual and reciprocal, forming an ecosystemic structure of continuous energy movement, release, exchange, reception and self regulation – a feedback system of crucial importance and great complexity. As said, in the days before the contemporary era these dynamics were largely unknown: space had a role which was very much formal and symbolic in nature, ignoring the compositional and conceptual possibilities, almost excluding them from the creative process. Many subsequent avantgarde and experimental art movements, inspired by the investigations of the new science of sound, tried to use its potentials in an active – more than passive - way: one of these movements, installation art, came to prominence with the aim of integrating the space in which the work resides as a part of the work itself. Modification of the way with which space is experienced, along with possible interaction with the fruitors, can then occur. Sound art is a consequential expansion of it: sound happens into space and therefore is used as an active part of the installation, dialoguing with it. Crucial, in this sense, is the addition – when compared to visual-only installation art – of the time element: sound evolves in time, and its evolution represents history and behaviour of the installation itself. In accord with these elements, it can be possible to trace and highlight the components and history of sound in space, through a systemic processing of its characteristics.

III.A SYSTEMIC PROCESSING OF SPACE: SYNTÁXIS

As said, space seen in the context of sound propagation – and under an ecological perspective - is a system of complex dynamics and relationships, where acoustic energy is produced, amplified or attenuated in its spectral characteristics, exchanged with the surrounding environment and recepted by the agents in it. Reaction can then occur, where the agents (e.g. living beings)produce sound in return and so on, in a feedback loop which is typical ecosystemic behaviour. An example worth of note is when exchange of spoken language happens: semantics aside, sound produced by a source propagates, is recepted and can provoke reaction with more production. The process of propagation, the second stage in this small but significant chain of events, carries with it a great amount of information: the energy moving through space is shaped and modelled by the particular physical characteristics of the environment – natural resonances, properties of absorption and reflection of the materials present in the environment, diffraction, and others - which constitute then a significant part of what is, in the end, percepted as sound. Consequently, it could be said that the whole moment of propagation, and the data contained thereof, represents the identity of the particular space in which sound occurs. An identity which the author's installation, *Syntáxis(Acoustic Generative Sound Processing System, part 1)* investigates in a number of different ways. Sound as element of space was already used as the main aim of the author's previous work of installation art, *5x4: a Neapolitan Soundscape*: in it, sound was extracted from his natural urban environment and presented for listening to several different people, thus exploring the different meanings of it when separated from context. Here, sound continues to be a material, but also comes into being as a way to structure an evolutionary discourse on itself, and on the history and peculiarities of the space where it propagates as well.

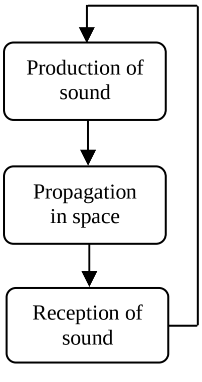


Fig. 1.Feedback loop of production – propagation – reception in an acoustic space.

The main backbone of *Syntáxis* is a systemic structure of sound reception, processing and reproduction, a chain composed of three elements.

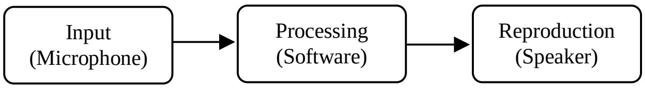


Fig. 2.Main structural design of *Syntáxis.*

Being an installation, *Syntáxis* occupies permanent space, working continously in its process. Sound happening in its surroundings is feeded in the system by means of a microphone, and processed by a software algorithm, developed on the Max/MSP 5 platform. The result then returns back into the installation's space, by means of a speaker (or more than one). An element of great importance here is that the installation has a cycle of processing which is carefully structured, and which aims at a progressive highlight of that number of informations which the sound material happening in the installation's space contains, and which are linked to it. The time element thus plays a role of great importance, also taking into account history of sound and therefore, by the installation's point of view, history of the space thereof.

IV.SYNTÁXIS' CYCLE: AN EVOLUTIONARY DISCOURSE

At the start of the installation's cycle, sound received by the microphone is first analysed by means of a Fast Fourier Transform on the signal received, and the two strongest peaks in the frequency spectrum are recorded. These peaks represent the peculiarities of the sound material, modelled by the natural characteristics of space in which it propagates, at the moment of the cycle's first stage. After a twenty seconds delay, *Syntáxis* starts an history: sound is processed by means of a bank of resonant bandpass filters, in couples, which represent a population. Twenty couples of filters, twenty fourth order bandpass filters that is, with their center frequencies equally positioned in the frequency spectrum, bandwidths large enough to permit a sufficient recognition of the sound processed. This array of individuals constitutes two "populations" of sound, each composed of ten individuals, and contains the potential that will be optimized by a genetic algorithm, a *search space* [II]aimed at its main objectives: the identification of the resonant peaks contained in the sound at the beginning of its history, the identity of space in a particular moment in time. It's a formalisation which takes in account, for its evolution, the characteristics of sound and acoustic space: the whole collective of sound data regarding the space where it is evolving. The key element here is fitness, the propensity for each individual – and thus the filter's center frequencies – of achieving their objectives.

Genetic algorithms generally emphasize – for their structural processing – the presence of a genotype, the representation of a single individual's genetic code, to be evaluated. This genotype is typically represented by a simple data structure, in most cases – as in *Syntáxis* – by a string of bits. The algorithm develops its optimization in a series of stages formally known as generations: in each generation, the population's genetic codes are reviewed and classified in a gerarchy based on fitness – typically an index ranging from 0 to 1 - towards a specific objective.

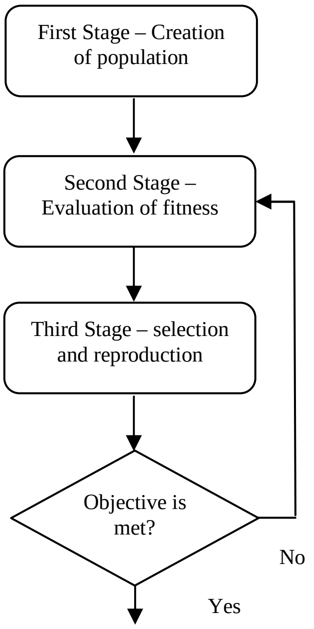

First Stage – Creation of population
Second Stage – Evaluation of fitness
Third Stage – selection and reproduction
Objective is met?
No
Yes

Fig. 3. Layout of a basic genetic algorithm.

As in the Darwinian model, survival of the fittest occurs, and individuals with greater fitness are selected for breeding and thus recombination of their genetic code: others with low fitness have greater chance of dying without reproduction. Offsprings are then inserted in the population, and the process starts again with another generation. History comes to an end when the objective is met. In *Syntáxis*, individuals – as said – are bandpass filter units, which act as instruments of investigation of sound, and genetic codes are represented by 32 bit strings equivalents of their center frequencies. At every generation, the individuals' genetic codes are compared with the resonant peaks analysed at the first stage of the installation cycle, which represent the objectives. Individuals with center frequencies which are closer to the objectives have an higher chance of recombining themselves, by mixing their genetic code together, into new individuals with better center frequencies. The more the filters go towards their objectives, the more the bandwidths get narrow, until – when objectives are met and the resonant peaks are all centered – they center their bandwidth on a single frequency, thus highlighting the peaks in their single uniqueness. When the average fitness of all the individuals of the population is high enough so that the filters center on the peaks-objectives, the process completes and restarts from the beginning, with another analysis.With this series of generations, the system recombines itself in time, as said. It thus traces its history, the history of sound in its space, using the time element as a background for its evolution, and investigates the properties that shaped the sound detected and analysed at the beginning of the installation cycle.

The sound material which *Syntáxis* processes is not fixed, though. Space is an entity with dynamics which are complex and variable: energy is produced, transferred into space, recepted and so on, as said. The installation applies its ongoing formalisation on an entity which is therefore dynamic, which changes, which is time-variant. An investigation of the product of space and its history in its ongoing, direct manifestations.

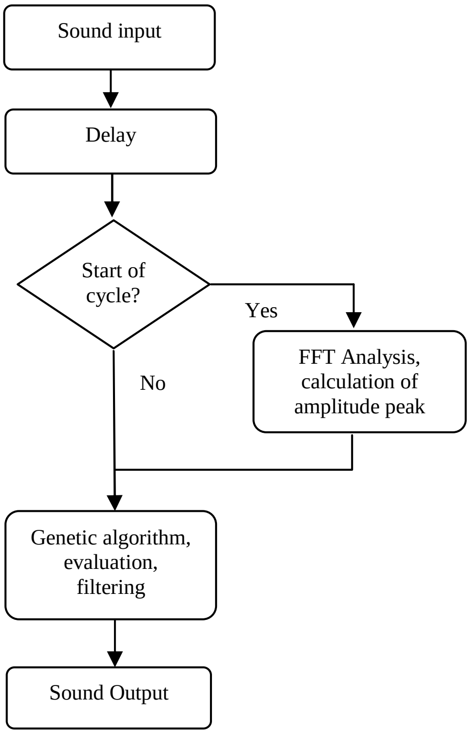

Sound input
Delay
Start of cycle?
Yes
No
FFT Analysis, calculation of amplitude peak
Genetic algorithm, evaluation, filtering
Sound Output

Fig. 4. Scheme of a single *Syntáxis* cycle.

It is worth noting that the *Syntáxis* chain implies a feedback loop. Sound is recepted from the installation's space, processed, diffused by speakers, and goes back into space, where it will eventually be recepted again along with new acoustic energy present in the environment. The presence of a twenty seconds delay line is justified by the obvious need to delay accumulation, thus setting up a self-gating system which will be briefly discussed – in the overall context of the Max/MSP 5 patch - in the next section.

V.A BRIEF LOOK AT SYNTÁXIS' MAIN SOFTWARE STRUCTURE

The core of *Syntáxis*' processing block consists of a Max/MSP 5.02 patch. It's worth noting that - since the genetic algorithm and the FFT analysis and peak calculations are based on a great number of iterations processes – for simplicity's sake the author has integrated the graphical object structure of Max/MSP with its javascript support. Thus, the whole DSP part is Max/MSP based, while the genetic algorithm and other elements are programmed in javascript and integrated in the patch. A thorough analysis of the software, in terms of specific references to code and the like, is way beyond the scope of this paper. But a couple of characteristics, vital to the overall design of *Syntáxis*, will be discussed.

It is of importance to note that *Syntáxis* comes in elements that could be called blocks. The chain of the installation has already been examined in this paper before: it consists of a line made by a microphone, a processing unit(i.e. DSP algorithm, discussed earlier) and a speaker.

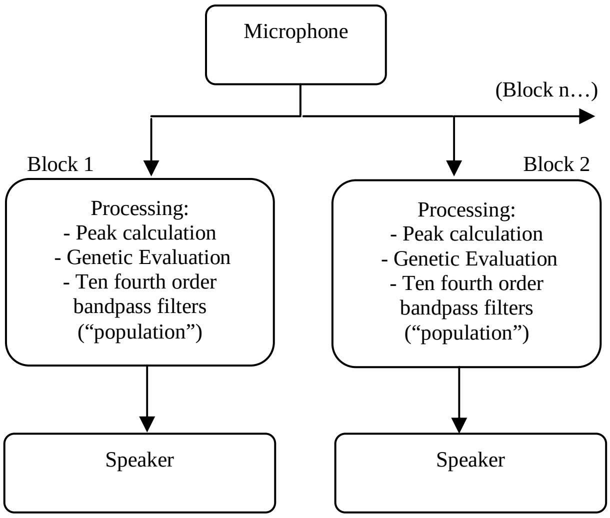


Fig. 5. *Syntáxis*' chain of processing blocks.

The processing and speaker elements constitute a single *Syntáxis* block. A single block has a population, array of fourth order resonant bandpass filters, of ten individuals: the whole process calculates a single peak-objective for its block, and genetic evaluation ensues. As of now, *Syntáxis* uses two speakers to reproduce back in its space its sound material, and therefore it means that it works on two blocks. That is, two populations, a total of twenty individuals (ten per population), a process of investigation of sound in space which will in time highlight the two strongest peaks in the spectrum analysed at the start of the cycle. This means that, in this way, we have two processes running at once, moving simultaneosly towards two different aspects of the same sound: two indipendently processed but correlated sound sources. Multiple blocks and thus multiple peaks can be processed at the same time, e.g. setting up a quadraphonic or octophonic configuration, therefore working on more peaks of the recepted/analysed sound.

A second element of note is feedback, which is consistent with the whole behaviour of *Syntáxis*, and on different levels, as said: symbolic, systemic, ecological, acoustic and, strictly speaking, audio. The microphone – speaker chain, coupled with the time element and the continous energy exchange, which is a central aspect of the installation, makes accumulation a possible and likely event, especially in case of slow evolution of a population. The dynamics of the installation space are also important: a great number of agents producing sound could minimize energy optimization in the environment, and thus negate feedback, while particular physical characteristics of the environmental space could instead strengthen certain frequencies more than others. A self – gating system, with the already mentioned delay unit, was created in the patch as a system of self regulation and response against saturation. The microphone input is multiplied by a gain factor, then goes through a twenty seconds delay line. Signal going out of the delay line is then both directed to the rest of the processing system and envelope followed. The resulting amplitude tracking , in inverse relation and as control signal, is low pass filtered with cutoff frequencies around 1 Hz (to smooth out ripples and low amp fluctuations) and sent back before the delay line, as factor of control/attenuation. The whole process carries with it the element of homeostasis, where the system regulates itself in order to maintain a stable condition[III]. – in this case, as said, to avoid saturation and consequently distortion[IV].

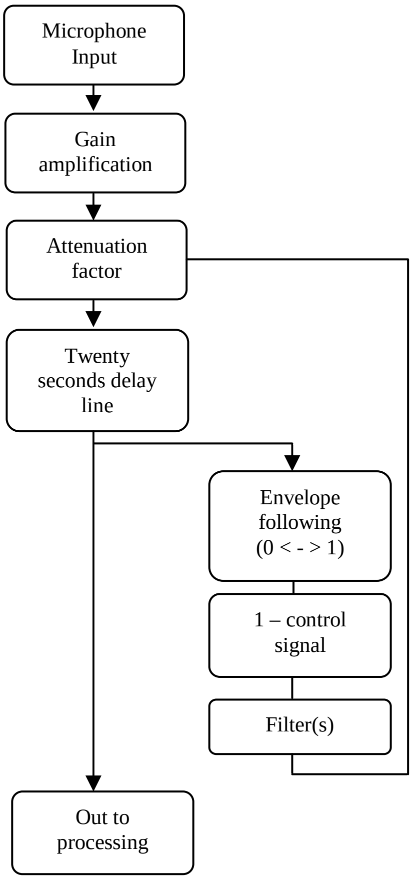


Fig. 6. *Syntáxis*' self-gating system.

VI. Conclusions, possible outcomes

Syntáxis is a work in progress. Its first incarnation, namely entitled *Part I*, explores only a small number of the whole possibilities of sound in its characteristics and relationships with acoustic space, although on a number of different levels. One of the main aspects of it is certainly investigation of space as a physical model of generating and shaping acoustic energy. Different, physical spaces can give very different results, not only through their dynamics of frequency amplification and attenuation but also through reflection, absorption and thus reverbation, and so on. More speakers could be used, thus having more peaks to be analysed and highlighted. Their placement in space could be adjusted in accord with specifics characteristics of the environment where the installation takes place. The process could be reversed, aiming at investigating frequency regions which are attenuated instead. Implications are numerous, and represent ispiration for future research and experimentation. Another aspect of *Syntáxis* which has been stressed a number of times in this paper is its recording and organic representation of history of sound in space. The concept of evolution in time has been represented in order to highlight this particular aspect. In the sense, the choice of using installation art to develop *Syntáxis* represent also the coherence of using space and sound in it, as a whole, in order to let it, in a way, to comment on itself.

ACKNOWLEDGMENTS

Special thanks, first and foremost, to Agostino di Scipio: his teachings, words and heredity were vital for the development of the many topics reported here. Thanks are due also to Marcel Wierckx for help and support with the very first, embrional stages of the patch.

REFERENCES

[1] Barry Truax – World Soundscape Project, *Handbook for Acoustic Ecology,* A.R.C. Publications, 1978

[2] Darrell Whitley, *An overview of evolutionary algorithms: practical issues and common pitfalls.* Computer Science Department, Colorado State University

[3] Agostino di Scipio, *Sound is the Interface: Sketches of a Constructivistic Ecosystemic view of Interactive Signal Processing.*CIMproc.,2003 http://xoomer.alice.it/adiscipi/CIM03b.pdf

[4] Agostino di Scipio, *Using PD for live interactions in sound. An exploratory approach.* 4th Linux Audio Conference proc., 2006

Textural Composition and its Space

Dr Kerry Hagan
University of Limerick, Limerick, Ireland, kerry.hagan@ul.ie

Abstract – The aesthetic implications of real-time stochastic sound mass composition call for a new approach to musical material and spatialization. One possibility is found in the fertile ground of musical texture. Texture exists between notions of the singular sound object and the plurality of a sound mass or lo-fi soundscape. Textural composition is the practice of elevating and exploring the intermediary position between the single and the plural while denying other musical attributes. The consequences of this aesthetic principle require a similarly intermediary spatialization conducive to textural listening. This spatialization exists between point-source diffusion and mimetic spatialization. Ultimately, the ramifications of textural composition affect both the space in the sound and the sound in space. This paper introduces the intermediary aesthetics of textural composition focusing on its spaces. It then describes an implementation in the author's work, *real-time tape music III*.

I.Introduction

Xenakis developed stochastic operations to create sound masses [1]. Since his work, composers have been exhausting computer methods for generating stochastic models. However, new development must continue in the application and aesthetics of stochastic methods, and not simply in the methods themselves.

Sound objects, Truax's "space in sound" (timbre), and "sound in space" (diffusion) [2] dominate the aesthetics of acousmatic music. Despite the generally, accepted fixed-medium definition, acousmatic music does not necessarily exclude real-time compositional practices [3]. Therefore, acousmatic music includes real-time computer-generated music provided it is preoccupied with sound objects, spatialization, and the sound image engendered by both.

The *real-time tape music* series evolved from stochastic sound mass composition contextualized by acousmatic principles of sound objects and space into something perceptually different: textural composition.

Textural composition resides in the intermediary regions between dialectical poles identified in electroacoustic, specifically acousmatic, composition. By subjugating gesture, primarily, and other musical attributes, secondarily, textural composition occupies the spaces between dialectics.

Dialectics in acousmatic music exist both in sound material and its spatialization. The single sound object opposes the lo-fi soundscape first described by Murray Schafer [4]. In spatialization, mimetic practices dissimulate the loudspeakers, while point-source compositions embed musical agency within the loudspeakers.

Spaces between dialectics can be experienced two ways. If gradient stages exist, an intermediary position can be a grey between black and white. If the dialectics are categorical, the intermediary position is marked by a fragile boundary where something can oscillate between each category. Music existing in the grey areas or on fragile boundaries begets intermediary aesthetics.

In acousmatic music, texture is poised on the delicate edge between the categorical singularity of the sound object and the ambiguous plurality of the soundscape. Spatialization in acousmatic music can exist in a shade of grey where mimetic aesthetics merge with the democracy of point-source diffusion. This intermingling creates immersive spatial motion in all perspectives without necessarily engendering psychoacoustically cohesive trajectories of sound in space.

First, this paper briefly contextualizes textural composition within acousmatic practice, then addresses the nature of texture itself in its duality with gesture. The focus on texture comes to the first dialectic: the sound object versus the soundscape. It proposes a sound meta-object, an intermediary between this dialectic.

The sound meta-object carries with it ramifications for both the space of the meta-object and the meta-object in space. This paper discusses space and the meta-object in terms of scale, movement, physicality of the listener, and perspective, arriving at the second dialectic: mimetic spatialization versus point-source diffusion.

Finally, a description of an implementation of textural composition, *real-time tape music III*, examining the stochastic and random processes used concludes the paper.

II.Arriving at Textural Composition

The aesthetic foundations for the sound-mass compositions leading to textural composition form the basis for the discussion of textural composition and the spaces it incurs. Therefore, a brief survey of the aesthetic questions contextualizes textural composition and aids in defining it.

The underlying purpose for the *real-time tape music* series grew from the need to dissolve popular distinctions in electroacoustic music. The aim was to pursue typically acousmatic objectives using traditional tape techniques while capitalizing on real-time computer-generated controls (e.g., stochastic processes) to create "live" acousmatic works.

One of the fundamental acousmatic concerns that contextualizes *real-time tape music* is its focus on space. Smalley states that "acousmatic music is the only sonic medium that concentrates on space and spatial experience as aesthetically central," even though he concedes that his ideas can be applied to "other electroacoustic music genres which possess an acousmatic component" including live and interactive genres [5].

Harrison makes a similar statement in comparison to instrumental music. Space can play a more defining role in electroacoustic music than instrumental music. However, Harrison emphasizes that this true in electroacoustic music especially "in which the cause of the sounds is not seen or necessarily implied [6]."

Therefore, the acousmatic sound sources (i.e., the veiled sound source) coupled with the central focus on space firmly situate *real-time tape music* as an acousmatic endeavor. Yet, Harrison's own definition of acousmatic music ("predominantly 'tape' music, music 'on a fixed medium' … descended from *musique concrète*") problematizes this context [6].

Harrison more profoundly segregates aesthetic approaches based on a much older distinction: that the remnants of the dispute between *musique concrète* and *elektronische Musik* could be found in what he classifies as "organic" and "architectonic" music. In architectonic music, structures depend on "quantifiable distances *between* musical events (in all parameters)." Organic music "explores the qualitative evolution" of musical events. And, all parameters include the spatialization of the events themselves [6].

Harrison explicitly defines sound diffusion as the "realtime (usually manual) control of the relative levels and spatial deployment *during performance*." And, he argues that the explicit structures of architectonic music requiring sounds to be at specific distances to each other at specific times do not lend themselves to such diffusion. On the other hand, live diffusion not only suits but supports the qualitative evolution of organic music [6].

Harrison's definition severely narrows the domain of spatialization. But even Smalley includes a similar definition as one aspect of what sound diffusion can mean: as "'sonorizing' the acoustic space and the enhancing of sound-shapes and structure in order to create a rewarding listening experience [7]."

Therefore, *real-time tape music* not only had to focus on space as its central issue, but it had to address real-time sound diffusion as well. Stochastic processes controlled the sampling and playback of sound files to create sound objects in real-time. These sound objects were then diffused in real-time using similar processes in order to suit the qualitative environment wrought by the objects.

The stochastic processes, however, embodied another aesthetic: the sound mass. Promptly, sound material and diffusion thickened until sound objects became sound masses. Powerful personal computing allowed for more layers of processes, objects, and spatialization. The result of this aggregation could only be heard as something significantly different than sound mass or sound objects. The work, *real-time tape music III*, explores this new domain, coined by Hagan in [8] as "textural composition."

III.Gesture and Texture

Many composers have identified a gesture/texture duality but often with different perspectives. Two perspectives, one by Murray Schafer and one by Smalley, offer a starting point for addressing the difference between textural composition and material that is textural.

Murray Schafer distinguishes gesture from texture in terms of number and attention. For Schafer, gesture indicates a unique event. Texture, on the other hand, consists of innumerable events. More important to this distinction is the perception of these polarities: gesture is the "noticeable," while texture can only be perceived "in masses or cluster formations [4]."

Smalley contributes another quality that divides gesture from texture: "The energy-motion trajectory of gesture is… not only the history of an individual event, but can also be an approach to the psychology of time [9]." Therefore, the real and subjunctive passage of time plays a fundamental part in whether material is gestural or textural.

Time as a function of music becomes forward motion or linearity. Smalley distinguishes gesture from texture in terms of this directionality. He states, "A music which is primarily textural, then, concentrates on internal activity at the expense of forward impetus [9]."

Murray Schafer and Smalley pinpoint the fundamental aspects of music that ultimately beget textural composition: number and its apperception, and a modified approach to time. These aspects lead to a dialectical investigation of the acousmatic sound image. And, it is in the boundaries between dialectics that textural composition lies.

IV.Sound object versus Lo-Fi Soundscape – Number and its Apperception

Descriptions of the metaphorical sound object are also metaphors of vision, taction, or corporality: volume, size, texture, mobility, etc. The intrinsic ambiguity of metaphors creates rich intermediary positions. The metaphor of texture offers a particularly fecund continuum. Textural composition is the practice of working within musical texture to subjugate other metaphorical qualities for the express purpose of creating a sonic space on the boundary between sound objects and the lo-fi soundscape defined by Murray Schafer [4].

A distinctive sound object exists because its spectromorphology distinguishes it from other sounds. Sound objects have sonic boundaries drawn by the time-varying frequency spectrum and gestalt behavior. In other words, the qualities of a sound object bound it away from other sound objects in the acousmatic image.

When a critical mass of sound objects comes together, the result can be the lo-fi soundscape, where objects are blurred into each other, or a sound mass, where objects are distinctive, but amount and behave together.

A multiplicity of sound objects without boundaries results in a lo-fi soundscape. In extreme cases, as Murray Schafer describes it, "individual acoustic signals are obscured in an overdense population of sounds. The pellucid sound—a footstep in the snow, a church bell across the valley or an animal scurrying in the brush—is masked by broad-band noise [4].” The notion of a soundscape still implies that, though indistinct, multiple sound objects exist.

The sound mass can be seen as an intermediary. It consists of individual objects, but takes on essential sonic boundaries much like the single sound object. Yet, at the core of the sound mass are individual units with their own boundaries.

Textural composition provides an alternative to the sound mass intermediary. One sound object expands to occupy the entire musical space, and its boundaries exist beyond the acousmatic image. Without boundaries or gestalt behaviors, it is not a sound mass. Though this could become a lo-fi soundscape, it is not. The imagination perceives the whole as one thing, not amalgamated multiple things. The sound object is magnified into a sound meta-object, and its musical attributes advance to the level of compositional material.

A sound meta-object requires substantive space, both in terms of the space in the meta-object and the meta-object in space. More importantly, it is crucial that the space facilitates texture’s dominance as the central musical material while diminishing the influence of other attributes.

V.Space in the Sound Meta-Object

Texture cannot exist alone; texture is a quality of some material. The rough grit of sandpaper requires the substratum of paper. Furthermore, the elements yielding the texture cannot be separated into individuated items without destroying the texture itself. To see or feel the individual grains of sand on sandpaper is to necessarily lose the “rough grit” experienced at normative perspectives. Yet, to discuss the texture of the sandpaper and nothing else means to frame only that aspect. One can say, “Feel the grittiness of this sandpaper,” and the attention is drawn to texture.

Drawing one’s attention to the texture of the sandpaper sacrifices the other qualities to which one can attend: e.g., the shape of the paper (is it a sheet or designed for a belt sander?), the color (e.g., is it emery or aluminum oxide?), or the flexibility (e.g., is it cloth or fiber sandpaper?).

Likewise, textural composition requires a shift in attention that diminishes other musical parameters. The sound meta-object serves this purpose. Several factors conduce to spawn a sound meta-object with sufficient space to carry texture: singularity, volume, and time.

First, the meta-object is a single unit. The edge between a unit object and multiple objects where the meta-object resides is categorical and requires a delicate balance of multiple factors. The tenuous existence of the singular sound meta-object depends on, among other things, the component parts’ mutual compatibility - that they appear to belong to one ideational thing. The grain of the texture must not stand out as single sound objects. Furthermore, these component parts must maintain an overall, unchanging, average spectromorphology to further enforce the singular.

Second, the volume of the meta-object must extend beyond the periphery of the acoustic “view” and subsume the listener, negating any potential for the whole to act as a sound mass.

Barrett identifies density, texture, and amplitude as key contributing factors to “implied spatial occupation [10].” Truax notes that spectral richness, duration, and asynchronicity collude to affect perceived volume [2]. On the one hand, Barrett is characterizing sound masses, while Truax is discussing a single, complex sound. But, both formulas apply to the sound meta-object.

However, these qualities of cyclopean volume contravene the characteristics of the singular meta-object. Therefore, textural composition exists on a categorical, fragile edge between the magnitudinous singularity and the multitudinous mass, an intermediary aesthetic [8].

Time becomes a factor as a result of the gesture/texture duality. As gesture is the opposite of texture, the temporal aspect of the meta-object must stretch to become the entire piece, providing no basis for the interaction of sound objects, linearity, or directionality. Therefore, gesture is neutralized, and time is subverted.

Smalley suggests that “high sustained, continuant morphologies” can suggest “space itself rather than anything which moves in it, something possibly atemporal, as if time is becalmed [5].” However, continuant morphologies typically have softer, if any, textures. On the other hand, the subjunctive space of the sound meta-object combined with its consistent overall spectromorphology tends to perform the same role after sufficient inculcation.

Though the meta-object subverts the time of the piece at a formal level, texture must still dominate at the local level. For texture to fully ascend to the role of main compositional material, texture itself must be developed. Music is a temporal art. Therefore, temporal aspects of texture must figure prominently to elevate texture’s status to the central musical material, even if the sound meta-object is timeless. Texture in a textural composition must be dynamic in order for it to be compelling.

VI.Sound Meta-Object in Space

Sound and its musical space is inextricably linked with its acoustic diffusion. The space of a sound meta-object carries with it implications for its existence in space through diffusion. In textural composition, this means that diffusion must correspond to the size and extension of the sound meta-object. And, the diffusion must complement the dynamism of the texture.

Additionally, the metaphor of texture in music derives from physical experiences of texture in vision and touch. Therefore, the spatiality of texture is predicated on the physicality of the human body. Any philosophical approach to spatialization needs to contend with space in relation to the listener.

A. Volume and Scale

The most concrete aspect of spatializing a sound meta-object applies to its perceived volume and scale. First of all, according to Truax, “Uncorrelated signals will increase the apparent volume of a sound provided there is a basis for perceptual fusion of these components into a single, possibly complex auditory image [2].”

Smalley argues that the distribution of spectral space contributes to volume and scale, as well. “I can create a more vivid sense of the physical volume of space by creating what I shall call *circumspectral spaces*, where the spectral space of what is perceived as a coherent or unified morphology is split and distributed spatially [5].”

In effect, assuming that (psychologically if not psychoacoustically) the sound meta-object is perceived as a unified morphology, then the components of the meta-object must be *uncorrelated* and *distributed* spatially. However, the components themselves are complex sounds. Therefore, creating uncorrelated copies of the constituent sounds and diffusing them throughout the space will increase the volume further still.

B. Dynamism and Motility

Spatial stasis does not crucially impact the perception of a dynamic texture, but motion does enhance a dynamic texture. As Smalley says, “The motion must be implicit in the sound itself or the texture itself or the context itself [7].” Since the texture of the sound meta-object changes in time, static diffusion of the texture seems incongruent or dissonant.

Mimetic spatialization fabricates a subjunctive space through psychoacoustically illusive or allusive localization and motion. The interaction of sound objects in a mimetic subjunctive space is called “objects-motion-environment” by Simon Emmerson [11], and it is a significant musical trope in acousmatic music.

However, if the spatial motion coheres elements of the texture into distinct sound objects through identifiable trajectories in space, the fragile meta-object is splintered into multiple sound objects. Therefore, diffusion needs to occur between stasis and mimetic spatialization.

C. Human Agent

On the most abstract level, perception of scale, motility, and mimetic trajectorial spatialization convolve in the human observer and ultimately rely on the physicality of the perception and the human perspective.

Texture is a physical quality, perceived through vision and touch. Smalley asserts that, as such, texture (among other things) has space because vision, touch, and sound “embody underlying spatial attributes.” More importantly, perception has physical roots in the body, “which is always at the focal centre of perception – as utterer, initiator and gestural agent, peripatetic participant, observer and auditor.” The importance of the human-centered perception means that musical perspective is always related to the human scale, what Smalley calls the “egocentric space [5].”

But, Smalley also suggests that gestures operate within the human scale because “if gestures are weak, if they become too stretched out in time, or if they become too slowly evolving, we lose the human physicality. We seem to cross a blurred border between events on a human scale and events on a more worldly, environmental scale [9].”

Since texture is the opposite of gesture, then textural composition, by its very nature, operates outside the human scale. This relationship to egocentric space only serves to strengthen the impression of volume, scale, and timelessness the meta-object strives to achieve. However, the spatialization of the meta-object cannot rely on the environmental qualities alone to elevate texture to textural composition. The perspective of the listener must be addressed.

D. Perspective

Perhaps the most useful discussion of perspective and the listener in the case of textural composition comes from Smalley’s space-form approach because it “places time at the service of space [5].” Since time is subjugated in the textural composition, the analysis should focus on aspects of space outside of time. For this reason, Smalley’s other writings on texture motion and gesture [9] have less relevance.

To summarize: Smalley defines perspectival space as the “relations of position, movement and scale among spectromorphologies, viewed from the listener’s vantage point.” Perspectival space includes the shifting perspectives of prospective space, panoramic space, and circumspace. Prospectival space and panoramic space contend with the frontal, forward view of the acousmatic image, while circumspace “encompasses the listener, with the possibility of approaching or passing over egocentric space from all directions [5].

Smalley tends to work primarily with the frontal image, since he works in more orthodox acousmatic methods [7]. However, his preference for the prospective space cannot serve the purposes of textural composition. Prospective space favors the ‘forward’ space; sounds occurring behind the observer are only significant in relation to the front. This perspective bounds the acoustic space. A sound meta-object cannot be bounded or limited, so prospective space cannot support it.

However, Smalley also defines a space with no favored orientation. He calls this immersive space, a kind of circumspace, “where the spectral and perspectival space is amply filled, surrounding egocentric space, where the pull of any one direction does not dominate too much, and where the listener gains from adopting, and is encouraged to adopt, different vantage points [5].”

Hence, textural composition must achieve immersive space, where the sound meta-object not only becomes greater in the perceived volume and scale due to its sonic characteristics, but also greater than the egocentric space and unlimited in its range around the listener.

More importantly, textural composition must take on what Smalley calls environmental dimensions because it implies spaces beyond the listening space [5]. Music that uses environmental sounds and causes them to interact in environmental dimensions dissolve boundaries because of the listener’s experience with the environment. Textural composition, however, does not enjoy the advantages of being environmental in this way. However, proper spatialization can recreate an environment, however fantastic, that the listener can experience.

In Murray Shafer's lo-fi soundscape, all perspective is lost and there is only presence. "The modern lo-fi soundscape possesses no perspective; rather, sounds massage the listener with continual presence. As the population of sounds in the world increases, soloistic gestures are replaced by aggregate textures. Textures and crowds are correlatives [4]."

At first glance, this appears to be the state to which textural composition must assay: no perspective, only presence. However, as Smalley points out: "high density is the enemy of low-level detail," and "a packed density of full spectral range … creates a solid wall" around the listener [9]. This perceptual wall around the listener will, at any distance, bound the space.

So textural composition must have some amount of perspective, if only to suggest that all perspectives – in all directions and at all distances – are completely included in the sound meta-object. Textural composition must have superperspective. In this instance, textural composition requires an intermediary aesthetic of spatialization.

E. Mimetic and Point-Source Spatialization

Mimetic spatialization dissimulates the loudspeaker to create a perception of space between speakers. However, most composers interested in imitating reality do so in order to simulate motion in space, or objects in a trajectory.

The trajectory-based aesthetic of mimetic spatialization enforces an orientation and perspective that undermines the creation of the immersive space required for the sound meta-object, since sound trajectories are more effective in the frontal image. The objects-motion-environment relationship of mimetic spatialization debilitates the sound meta-object, shattering it into disparate elements.

An alternative, point-source spatialization, empowers each loudspeaker with its own musical presence. Different methods of point-source spatialization still allow the loudspeakers to retain their agency as musical performers, as Burns discovered in several of his works [12]. The democracy of the point-source method ensures that no one perspective is favored.

Yet these electromechanical performers are static. Any material that passes between speakers moves much like material would within an orchestra. The motion is only approximate.

Textural composition requires plenary motion, typically associated with trajectory-based mimetic spatial aesthetics, while employing an application that favors no singular perspective. The spectrum between mimetic and point-source spatialization is a gradient, so there are many degrees to which something may be more or less trajectorial, or more or less point-source. In this sense, textural composition occupies only one possibility within an intermediary aesthetic.

VII.Aesthetic Conclusions

Given the context of "live" acousmatic music, textural composition requires real-time generation of material diffused in space. Space, both space in the sound and the spatialization of sound, needs to figure prominently in the compositional method.

Textural composition is a practice of intermediary aesthetics, lying between dialectical poles. In sound material, a textural composition exists on the fragile categorical boundary between sound object(s) and the lo-fi soundscape. It exists in a sound meta-object, a massive, singular sound object that extends beyond the periphery, both imaginatively and in spatial diffusion. Furthermore, texture must ascend in dominance over other musical traits, and temporal flux of texture secures that position.

Likewise, the spatialization of textural composition must support the delicate existence of the sound meta-object. It must augment the volume and scale, it must not dissolve the meta-object into multiple sound objects, and it must match the dynamism of the texture with correlative motility. These qualities suggest an intermediary spatialization in a gradient between mimetic, trajectory-based aesthetics and the immersive space suggested by point-source aesthetics.

VIII.Compositional Implementation

In this paper, a stochastic process is a goal-oriented random process that has an equilibrium state as originally defined by Xenakis [1]. A random process is any probabilistic operation. Textural composition arose from developing stochastic and random techniques for the generation of sound masses using Pure Data (Pd) by Miller Puckette.

Random and stochastic processes only require a moderate number of parameters to affect drastic changes in sonic output, a great advantage in creating musical textures. Processes that control actions on the structural level provide the composer with the freedom to choose sounds sources that work together in a texture.

The author's work, *real-time tape music III*, moves away from previous sound mass composition and utilizes real-time random and stochastic processes at micro- and mid-levels to create a textural composition. Gaussian processes sample sound files and apply common tape-based manipulations to generate individual sonic events. Markovian stochastic processes determine the overall textural shifts. Uniform random processes control spatialization. The macro-structure, i.e., form, is predetermined.

A. Form

The piece consists of two contiguous movements. The first movement initiates the listener to textural listening through a didactic accumulation of sound objects. A crescendo finishes the first movement after approximately five minutes. The opening of the second movement rebuilds the texture for approximately one minute, leading to the main section of the second movement.

The main section comprises the majority of the piece, lasting approximately ten minutes. It is within the main section that the piece truly exhibits textural composition. The final thirty seconds of the piece is a slow diminution to silence.

B. Random and Stochastic Processes in Texture

Processes originally created for *real-time tape music I* and *II* generate individual texture streams from ten separate soundfiles. Gaussian random number generators take in mean and variance values for the playback speed, loudness, duration of the sample, and onset time within the soundfile. A sample is triggered in uniformly random increments of time. The samples are layered to create a single audio stream of an individual texture.

The parameters are set at the beginning of the first movement. Then, they change at the beginning of the second movement. Through out the duration of the second movement, the parameters ramp to new values by the middle of the second movement, returning to the initial second movement values by the end. This creates a subtle textural trend over the course of the main section.

The sound sources were chosen for their ability to ally with each other to create a meta-object, but with enough distinguishing features to create rich and diverse textures. Sources include a close-mic'd elastic band, close-mic'd carbonation bubbles in an aluminum can, a solo cello musical passage, a musical passage for cello and percussion, orchestral attacks, a plucked aluminum tab, a digitally processed musical passage for violin solo, processed samples of violin crunch bowing and harmonics, and a musical passage for woodwinds.

Each soundfile creates a single texture stream (see Fig. 1). Eight combinations of individual texture streams were chosen for their textural interest (see Fig. 2).

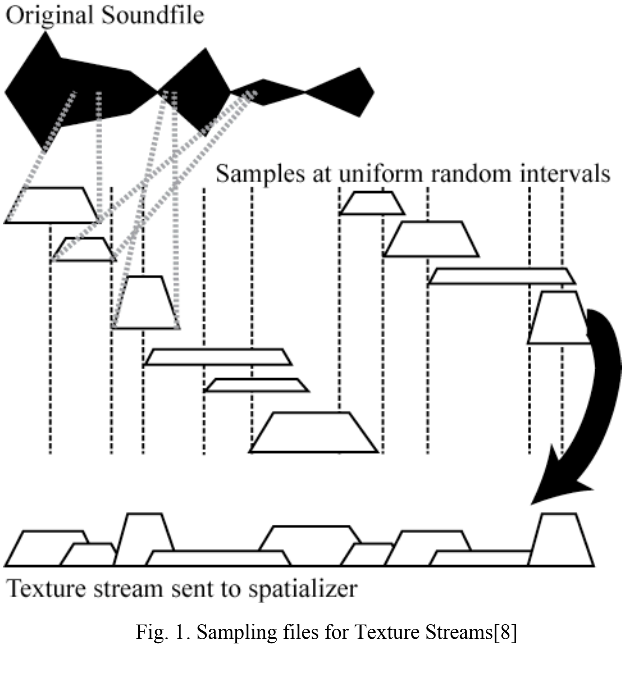

Original Soundfile
Samples at uniform random intervals
Texture stream sent to spatializer

Fig. 1. Sampling files for Texture Streams[8]

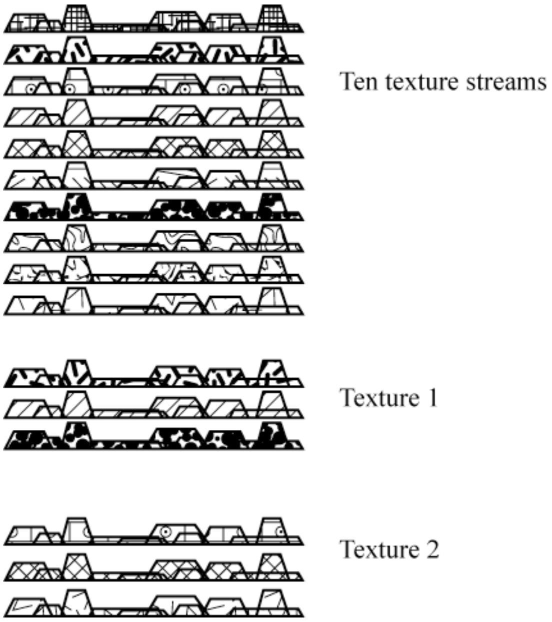


Fig. 2. Textures created from Texture Streams[8]

A Markovian stochastic process first invented by Xenakis for *Analogique A* and *B* determines the order and choice of texture. A portion of the patch used to analyze *Analogique B* is used for the selection of textures in real-time [13]. The pertinent quality of the Markovian process is that it is stochastic, i.e., it has an equilibrium state. At specific times in the second movement, the system is "seeded" with one texture for thirty half-second cycles, or fifteen seconds. The system then nears its equilibrium state with each fifteen-second cycle.

C. Spatialization

The processes controlling the random sampling and texture choices are computationally expensive. Therefore, the spatialization must rely on minimal resources. Loudspeaker amplitudes, interaural time differences, and artificial reverberation with uniform random variables create the motile environment without spatializing each sample.

Spatialization is mapped in the space as a circle. Each individual texture stream is given a constantly changing virtual angle that moves around the circle at varying speeds.

A simple equation calculates the relative amplitude of the texture stream based on the angle of the speaker in the circle and the virtual angle of the stream. This sends the texture stream in a path around the circle (Fig. 3 A).

A copy of the texture stream is placed 180° across the space. The copy is slowly oscillated in and out of phase with the original stream by virtue of a variable delay. The result of this creates the perception that the stream is crossing the space as well circling it. This also creates an uncorrelated copy of the stream, which increases perceived volume, as well (Fig. 3 B).

Finally, the stream is sent to a reverberation patch, and the amount of reverberation constantly varies, controlled by a uniform random number generator. This variable reverberation fabricates a more environmental feel, extending the volume of the sounds into larger imagined spaces (Fig. 3 C).

Since the individual texture streams are being spatialized independently, any given texture can have up to three separate spatialized streams. Consequently, an expanse is created where no single psychoacoustic trajectory sweeps the space, but rather the impression of frantic, turbulent motion pervades the space (Fig. 4).

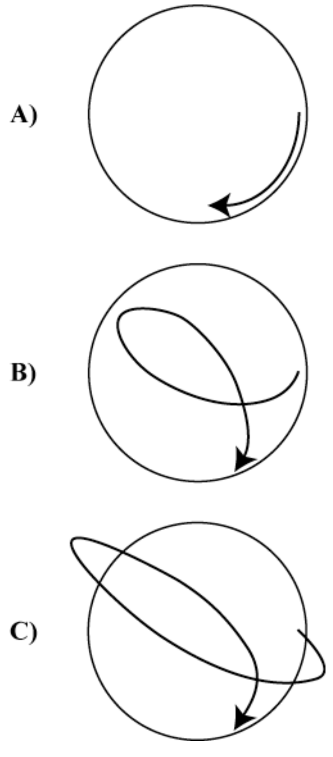


Fig. 3. Effects of Spatialization [8]

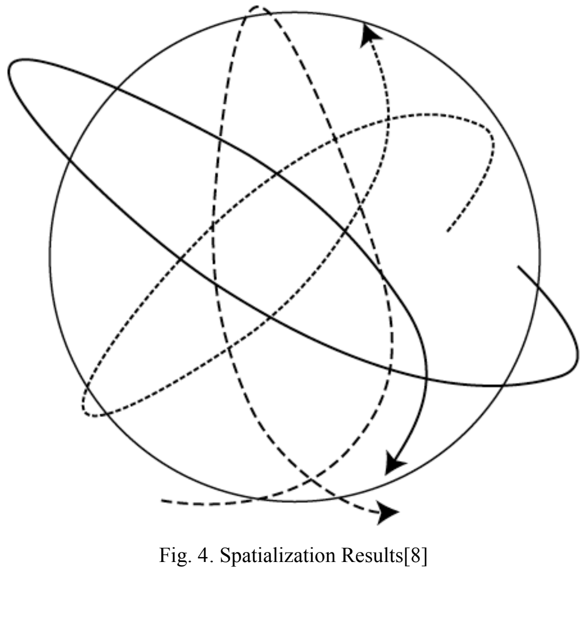


Fig. 4. Spatialization Results[8]

IX.CONCLUSION

Experimentation with stochastic and random processes creating sound mass compositions led to a new approach to conceiving a work, especially within the aesthetics of acousmatic music.

Textural composition, an intermediary aesthetic between sound object(s) and soundscape, wrestles with questions of gesture, number, volume, scale, and time in the instantiation of a sound meta-object. Congruent spatialization mediating between mimetic aesthetics and point-source principles augment the sense of the sound meta-object through extensive mobility, the physicality of space, and superperspective.

One realization of textural composition is *real-time tape music III*, a real-time computer-generated work. This piece utilizes Gaussian and uniform random variable generators for micro-level events and spatialization, while Markovian stochastic processes control structural events.

Due to the computation required for the random processes, an efficacious spatialization environment is created using loudspeaker amplitudes, interaural time differences, and direct-to-reverberant sound ratios to create plenary motion with little psychoacoustic coherence into trajectories.

REFERENCES

[1] Xenakis, I., *Formalized Music: Thought and Mathematics in Composition*. Revised ed. Harmonologia Series No. 6. 1992, Stuyvesant, NY: Pendragon Press.

[2] Truax, B., *Composition and diffusion: space in sound in space.* Organised Sound, 1998. **3**(2): p. 141-6.

[3] Wishart, T., *Audible Design: A Plain and Easy Introduction to Practical Sound Composition*. 1994: Orpheus the Pantomime.

[4] Murray Schafer, R., *Our Sonic Environment and the Soundscape: the Tuning of the World.* 1994, Rochester, Vermont: Destiny Books.

[5] Smalley, D., *Space-form and the acousmatic image.* Organised Sound, 2007. **12**(1): p. 35-58.

[6] Harrison, J., *Sound, space, sculpture: some thoughts on the 'what', 'how' and 'why' of sound diffusion.* Organised Sound, 1998. **3**(2): p. 117-27.

[7] Austin, L., *Sound Diffusion in Composition and Performance: An Interview with Denis Smalley.* Computer Music Journal, 2000. **24**(2): p. 10-21.

[8] Hagan, K. *Textural Composition: Implementation of an Intermediary Aesthetic*. in *International Computer Music Conference*. 2008. Belfast, Ireland.

[9] Smalley, D., *Spectromorphology: explaining sound-shapes.* Organised Sound, 1997. **2**(2): p. 107-126.

[10] Barrett, N., *Spatio-musical composition strategies.* Organised Sound, 2002. **7**(3): p. 313-323.

[11] Emmerson, S., *Aural landscape: musical space.* Organised Sound, 1998. **3**(2): p. 135-40.

[12] Burns, C. *Compositional Strategies for Point-Source Spatialization*. eContact! **8**, 2, http://cec.concordia.ca/econtact/8_3/burns.html

[13] Hagan, K., *Genetic Analysis of Xenakis' Analogique B*, in *Electroacoustic Music Society*. 2005: Montreal, Canada. http://www.ems-network.org/spip.php?article150

Sound in Cyberspace: Exploring Material Technoculture

Marilou Polymeropoulou

National and Kapodistrian University of Athens, Dept. of Music Studies/Media Studies, Athens, Greece,
m.polymeropoulou@gmail.com

***Abstract* — Cyberspace is nowadays a social network of people that produce, reproduce and consume technology culture, or as it is better expressed, technoculture. In this vast environment, transmittable digital information represents sound. However, what is the function of sound and why does it matter? In the following pages, I shall present sound as the materiality of technoculture in cyberspace, or, the cultural meanings of sound beyond natural space.**

INTRODUCTION

Almost at the end of the first decade of the 21st century, it appears that sound and space's materialities are still being reformed. Twenty four years after Gibson's novel, *The Neuromancer*, technoculture is the sort of culture that defined our age by making it digital [22]. Extending the idea of space to a non-natural place, cyberspace is where new culture resides; on the Internet – or the World Wide Web –, in science fiction and virtual reality, information is the vessel of sound [19]. Even though many theorists deny the use of word 'cyberspace' for the Internet, I will use it as a metaphore. I shall present sound and music in cyberspace through two cases: commodification on the internet and soundscape in cyberspace. These specific examples are used as they are characteristic of their symbiotic relationship – as will be analyzed furthermore. In this paper, I examine the material technoculture that is expressed through the cultural meaning of sound in cyberspace.

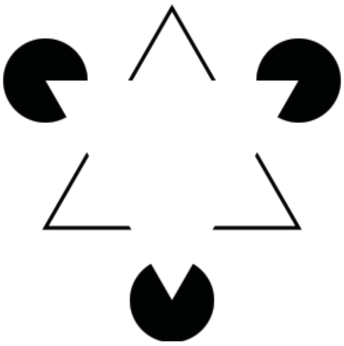


Fig. 1. Cyberspace's Symbol

I. THE NOTION OF MATERIALITY IN CYBERSPACE

It seems hard to explain music and sound through a material culture theorem. In philosophy, music's materiality is often discussed: sound is definitely not an object, but matters to us as an intagible artifact. On the other hand, music consists of particles of sound, which are immaterial airwaves that metaphorically derive their materiality. Futhermore, who would doubt the materiality of the harmonic structure of a musical piece? Very certain chords, in very specific places construct a very unique musical piece, and this only. It cannot be mistaken for any other if expressed through this structure. To expand this idea of materiality, music may not be itself material, but it extends its materiality to tangible music related objects, for instance, a compact disc, a musical instrument, or even, a speaker. Moreover, music in cyberspace depends on interaction and is based upon relations (composer-work, audience-work, performer-composer etc) [17].

Keeping the above thoughts in mind, I sum up that what actually characterizes the material culture of music in cyberspace is its *agency* [11]. Material culture represents the things that not only have matter, but also matter to people, and can be explored by analyzing relationships between people and objects [18, 23]. In music, one can explore the relationships between the composer and its musical piece; the musician and their musical instrument; the audience and the musical piece. Through these relationships, one can also understand the composer-audience relationship. "Things", argues Gell, "can appear as 'agents' in particular social situations; and so […] can 'works of art' [11]. Music, as a work of art on its own turn, acts like an 'agent', constructing a *Sonic Art Nexus*. This nexus can be used to recognize, analyze and understand the agents' relationships mentioned above.

Advancing the idea of a Sonic Art Nexus within cyberspace, agency becomes digital. All data –whether it is music, image, software– is translated to bits and bytes, computer languages and codes. Cyberspace has more agents to contribute that mainly act as the medium. A computer user wants to listen to a musical piece in mp3 form. What they have to do, is put the song into a specific music program that functions accordingly to a specially programmed software. The computer reads the translated information from the program and hence, produces a specific coded information that represents the song's digital structure. Consequently, data translates into sounds that are produced – with the aid of a soundcard (harware) – through speakers which are attached to the computer. All the encoded data represents a human-machine way of communicating, without which communication would not be possible. To sum up, cyberspace is the place where "transformation of modes of communication and information" occur [21].

In cyberspace we can examine social relationships between all kinds of agents. There are artists that promote their work, users that consume it, all sort of cyborgs appear ready to interact. What is of great importance is to research and analyze the role and/or function of sound and music in a non-natural, virtual place whose materiality is embodied in its users.

II. SOUND IN CYBERSPACE

In Attali's Noise: The Political Economy of Music, the writer makes it clear that "all music, any organization of sounds is then a tool for the creation or consolidation of a community, of a totality. It is what links a power center to its subjects, and thus, more generally, it is an attribute of power in all of its forms. Therefore, any theory of power must include a theory of the localization of noise and its endownment with form." [2] Music and sound could not be absent in cyberspace, "the electronic meeting place where disembodied communication takes place"[2] It is a virtual-place metaphor [1, 14] of a reproducible world that virtually exists and matters. The internet functions as the backbone of cyberspace. It serves as the main connecting point for many other networks [15]. In Baudrillard's terms, "the emergence of the internet as a kind of cybernetic terrain marks the end of the symbolic distance between the metaphoric and the real"[13].

Musical experience within cyberspace stresses the question of kind of bodies that do and do not appear in virtual worlds [13]. In other words, the appearance and interactivity of virtual-bodies enforces a re-examination of the body both as a physiological entity and phenomenological experience [13]. Accordingly, virtual bodies tend to "ignore or denigrate the dynamic and sensory capacity of materiality, both in the world and in our own bodies" [3]. Music in cyberspace demonstrates all the basic interactive characteristics of technoculture [19]: it is sensed through the natural body of the netter and hence, reacts virtually through their phenomenological avatar. In the following part, I will analyze music and sound materiality in video games that take place in cyberspace.

In the following part, I will describe two categories of sound functions that I will argue in my paper. Firstly, sound as a commodity in cyberspace, secondly sound as cybersoundscape. These represent two of the most common and characteristic uses of sound and music within what we call technoculture.

A. Sound as a commodity.

In our time materialities are constantly transforming: new technologies, new materials, even new ideologies have the tendency to alter needs and desires. Leaving partially behind the 'hardware' period of humanity, 'softwerism' seems to lead forward. In music composition the phenomenon is not new; electronic music is more than fifty years old and always advancing. Sound and music's so-called 'stylistic morphing' becomes even more apparent for non-specialists, that is, pure users of the cyberspace world. In Beer and Sandywell's words, music culture became digital and is thus characterized by the reinsrciption of a musical system or elements of that system in digital form, the construction of a continuum of elements and the application of transformation rules [4]. Music represents a commodified product multiply distributed in cyberspace [22]. However, the impact of 'music digitization' in cyberspace is firmly related to immaterial consumption and other marketing possibilities of virtual commerce[22].

Internet capitalism became even stronger with consumers' online shopping experience [7]. Shoppers often prefer Internet shopping due to its convenience and lower prices – which can be easily compared with few clicks [8]. As it was mentioned before, cyberspace includes metaphorically the virtual places of Internet and the World Wide Web [1]. Any shopper can visit any virtual music space (e.g. Amazon, HMV, Virgin Megastore, iTunes or any other online music related shop) in order to listen to samples of music CDs and moreover, with few more clicks, buy the CD – and at several times, any other music merchandizing, from collectable bootleg recordings to special limited edition CDs that were made available for specific markets in different countries around the world.

Music as a commodity usually brings about copyright issues. There is a whole range of internet websites offered that one can exchange music files with other users of the Web. Using different protocols, users have the opportunity to download pieces of music encoded in the popular compressed MP3 format legally or illegally [10]. From Napster and Kazaa download software to torrent clients and rapidshare, music was and is available online but such a consumption is considered illegal as these digital music files which anyone can upload on the Internet, are not protected by copyright law.

Commodification of music in cyberspace often becomes apparent in a variety of virtual communities. Online meeting places for musicians such as Noiz, Pro Session Music and Indaba Music attract musicians around the world who have the opportunity to share their musics with each other and exchange thoughts and ideas through forums. In certain cases, musicians also collaborate in order to produce a collective piece of music, a hybrid or in other terms, a cyborg musical piece.

B. *Sound as Cybersoundscape in Video-Games*

As a cybersoundscape I consider any background music. Firstly, that is noticeable in video-games. In our days, three-dimension online games have gained popularity. Many netters around the world experience cyberspace through MMORPG (Massive Multiplayer Online Role Playing Games) such as World of Warcraft, Second Life, Crisis, Line Age etc. Before these, MUD (multi-user dimension) games and MOO (multi-user dimension, object oriented, subspecies of programs known as MUD) were classes of programs that distinguished “interactive fantasy game which simulates a terrain through textual descriptions” [19]. Users became “characters in a world where they can interact with their environment and, most imporantly, with other players” [19]. Audio effects and music though, were part of all these game categories, and their role in video-games can be traced in the early ‘70s.

Early video-games’s music was stored on a physical medium such as the cassette or phonograph records. As technology advanced – and costs of materials were going low – programmers developed codes for music applied in games. From 8-bit music programming, to recent personification of sound, video-game music has been an integral part of the foreground [24]. As Ian Wall, video-game music composer states, “playing all those arcade games, I never even paid attention to the music. It just sounded like sounds to me. However, you know all the tunes. It’s so funny. The bleeps and bloops, they kind of invade your brain” [24]. To this, Tommy Tallarico, video-game composer adds: “if you remember in Space Invaders, you know, as the ships started to come down, the aliens, and as they got closer and closer, the sound got faster and faster. Now, what the game programmers did was that they took the person’s heart rate, and as they’re getting closer and closer, people would start to panic. Now they’d do the same studies without the sound, and the people wouldn’t panic as much. And it goes to show and prove how significant audio and music are”[24].

In other adventure games, which require puzzle solving skills and patience, such as the cases of *The Legend of Zelda* and *Grim Fandago*, Collins states that sound is critical in helping the player to adapt to the game issues. [5] Collins sees three types of sound in games: Interactive audio (when the player pushes a button and a sound effect is played for a specific action), adaptive audio (sound changes according to game environment, without a response to the player or listener), and dynamic audio (a combination of the above) [5].

To this, one can add up music’s function as a method of promotion. Collins underlines the symbiotic relationship between the music and games industry and the way popular music is used to promote video games and vice versa. [6] There are three categories of video games constructed upon music: musician-themed games (*PaRappa the Rapper*, SCEI 1996), creative games, (*Guitar Hero*, Red Octane 2005) and rhytm-action games (*Amplitude*, Harmonix 2003). [6]

There is plenty more information that can be added in the relationship of music, sound and cyberspace. What is of great importance remains the fact that one needs to understand the facts and be ready to reply to aesthetic/philosophical questions that they will draw.

III) Conclusions

Every code of music is rooted in the ideologies and technologies of its age [2].From the two paradigms mentioned, several conclusions can be made. What is mostly promoted through this perspective is the idea of music/sound as a commodification in a non-natural space. In a sense, hyper-commodification of popular culture can examine where fetishism of technique meets commodity fetishism. What really matters here, is to focus on the results of this phenomenon and the changes of the technoscape [19]. Cyberspace is often seen as an allegory to Plato’s Cave: it appears as an immaterial place that exists in the sphere of philosophy. Cyberspace is more than “a consensual hallucination experienced daily by billions” as proposed by Gibson. However, cyberculture becomes part of our natural life embodying the post-human notion of man, that extends their capacities. What is left to be examined, is how does that affect our relationship to sound and music? And even further, music as a cybersoundscape is it a selection of sounds without an aesthetic result?

References

[1] P. Adams, “Cyberspace and Virtual Places”. *Geographical Review*, Vol. 87, No. 2, pp. 155-171, 1997

[2] J. Attali, *Noise: The Political Economy of Music*. Manchester: Manchester University Press, , 1985

[3] B. Becker, “Cyborgs, Agents, and Transhumanists: Crossing Traditional Borders of Body and Identity in the Context of New Technology”. *Leonardo*, Vol. 33, No. 5, Eighth New York Digital Salon, (2000), pp. 361-365

[4] D. Beer, B. Sandywell, “Stylistic Morphing: Notes on the Digitisation of Contemporary Music Culture.” *Convergence* 2005; 11; 106

[5] K. Collins, “An Introduction to the Participatory and Non-Linear Aspects of Video Games Audio”, *Essays on Sound and Vision*. Helsinki : Helsinki University Press, 2007, http://www.gamessound.com/texts/interactive.pdf

[6] K. Collins, “Grand Theft Audio? Popular Music and Intellectual Property in Video Games”. *Music and the Moving Image*, Vol 1/1, University of Illinois Press, 2008, http://www.gamessound.com/texts/collinsGTA.pdf

[7] R. R. Dholakia, K.-P. Chiang, “Shoppers in Cyberspace: Are They From Venus or Mars and Does It Matter?” *Journal of Consumer Psychology*, *13*(1&2), 171–176, 2003

[8] R. R. Dholakia, K.-P. Chiang, “Factors Driving Consumer Intention to Shop Online: An Empirical Investigation”. *Journal of Consumer Psychology*, *13*(1&2), 177–183, 2003

[9] C. Eve Matrix, *Cyberpop: digital lifestyles and commodity culture*. London: Routledge, 2006

[10] Michelle W. L. Fong, “Music in Cyberspace”. Issues in Informing Science and Information Technology

[11] A. Gell, *Art and Agency*. Oxford: Clarendon Press, 1998

[12] W. Gibson, *The Neuromancer.* 1984

[13] D. Holmes, *Virtual Politics: Identity and Community in Cyberspace*. London: Sage, 1997

[14] D. Hunter, “Cyberspace as Place and the Tragedy of the Digital Anticommons”, *California Law Review*, Vol. 91, No. 2, (Mar., 2003), pp. 439-519, 2003

[15] S. Jones, *Cybersociety 2.0: Revisiting Computer-Mediated Communication and Community*. London: Sage, 1998

[16] S. Jones (ed), *Virtual Culture: Identity & Communication in Cybersociety*. London: SAGE Publications, 2002

[17] V. Mikić, “Technoculture: Subjectivity in the Net of Music”. *Spacesofidentity 4.1 (2004)*, pp. 73-82

[18] D. Miller, *Materiality*. London: Duke University Press, 2005

[19] K. Robbins, F. Webster, *Times of the technoculture: from the information society to the virtual life*. London: Routledge, 1999

[20] A. Ross, R. Lysloff, L. Gay, *Music and technoculture*, Connecticut: Wesleyan University Press, 2003

[21] R. Shields, *Cultures of internet: Virtual Places, Real Histories, Living Bodies*. London: Sage Publications, 1996

[22] T. Terranova, *Network Culture: Politics for the information age*. London: Pluto Press, 2004

[23] I. Woodward, *Understanding Material Culture*. London: Sage Publications, 2007

[24] “The evolution of video game music”, Article from http://www.npr.org/templates/story/story.php?storyId=89565567 April 2008

Towards a decodification of the graphical scores of Anestis Logothetis (1921-1994) . The graphical space of *Odysee(1963).*

Baveli Maria-Dimitra,
Music Department, University of Athens,
Hochschule für Musik "Franz Liszt" Weimar, maria-dimitra.baveli@hfm.uni-weimar.de

Georgaki Anastasia,
Music Department, University of Athens, georgaki@music.uoa.gr

Abstract — **In this presentation we are going to examine, via de-codification of graphic scores, the work of the avant-garde composer and pre-media artist Anestis Logothetis, who is considered one of the most prominent figures in graphic musical notation. In the primary stage of our research, we have studied these graphical scores in order to make a first taxonomy of his graphical language and to present the main syntax of his graphic notation, aiming at a future sonic representation of his scores. We also present an example of graphical space through his ballet *Odysee* (1963).**

Introduction

In the late '40s – after the 2nd World war –musical practice, following the developments in the visual arts, adopted a renegade stance moving away from the musical tradition that was prevalent up to that point. The palette of sound possibilities was radically enriched – incorporating electronic sounds and noise – and the new theories and techniques were followed by equally radical changes in musical notation. Composers had created their own private, individual way of notating their thoughts, and a new interest was developing among this generation of experimentalists in the organizational potential of each of the parameters of sound. From 1951, the role of the performer and listener became increasingly central to the conception of several works. Some of these new approaches were received as functional and essential for the evolution of music as practice; others though were criticized as being extreme and exaggerated.

The illustrations – the graphic pictures used as scores – show the extent of freedom and responsibility given to performers and reflect the emergence of chance as a defining element of this novel musical practice. The resemblance of some of the scores to the work of contemporary painters, such as Modrian, Mirò and Klee is striking. This growing interest in the visual qualities of graphic (pictorial) notation is a fact. Numerous collections of such scores have been arranged for exhibition in museums and galleries. There was an emerging feeling that graphic scores, independent of their function as notation, also have a meaning in the visual realm. The exhibition of those scores brought forth a fundamental question: "how can graphic musical scores stand as a form of visual art?" Pictorial notation can be perceived as art in the sense that it involves the drawing of symbols. While to the composer the use of a pictorial or graphic notation is strictly functional, to the viewer it is a drawing to be interpreted as a form of visual art.

1. The polymorphism of Logothetis' notation

Logothetis's urge to start a new, different musical notation derived from the problems music was facing during his time. According to his words: *In this way, the compass of modes of musical expression is significantly expanding. However, the notation with which one wants to describe the sonic events is not adequate for this purpose. Several problems arise out of this: not only noise-like sounds can only be represented with great effort or in some cases not at all, but the desire for a flowing music whose genesis can be experienced again and again is impossible to realize in this way*[1]. Considering therefore that the musical material was infused with new sonic possibilities, Logothetis tried to find a way to include into notation some additional parameters of sound, such as the time parameter of the musical structure, the positioning of sound in space, timber that can attain the quality of noise, and the homogenized flow of sound masses. Because of these additional parameters, this graphical notation system was more flexible and polymorphic than the traditional one. In addition, the composer tried in some way to protect on one hand the uniqueness of each performance and on the other hand the preservation of a piece's principles, giving the musicians designated flexibility (while performing). According to Logothetis: *What fundamentally differentiates graphic notation from traditional notation is the afore mentioned polymorphism, which clearly enables all performers to retain their subjective reaction times. The composer takes into consideration the divergences of the different performers in composing and expects a certain degree of surprise through the new formalization of musical form in every performance*[2].

[1] Anestis Logothetis, *Kurze Musikalische Spurenkunde. Eine Darstellung des Klanges*, Melos, 1970, Bd. 2, p. 39-43

According to Logothetis the polymorphism of graphical notation has both to do with space and with the method by which it is read. *Traditional notation is divided into systems and is read from left to right, like books. But since sound does not behave in the way written word does, we could think about using pictorial notation to represent musical events. (...)because musical time doesn't follow any direction, let alone the conventional left to right writing found in literary forms*[3]. The duration, time in a musical piece, is associated with motion in space and determined by its placement within it. *When a piece of paper is used as a space for representing sonic events, every point and line is brought into relationship with the entire surface and is temporally defined as short or long. This arbitrary correspondence between surface and symbols allows for the temporal associations of sonic events and the control of their duration; while out of convention, the positioning of musical events high or low on the paper represents high and low pitches respectively.*[4]

During the development of his notation system in the 60s, Logothetis completed more than 100 graphical notations.

At this point we should explain the terms "graphical notation" and "musical graphic", that were coined in an attempt to define this new script in music. *Notations are, generally, systems of signs/symbols, but the musical graphic is a painting, a drawing*[5]. The important difference between them is the fact that a graphic, though musical, *conveys a meaning not because of its use of signs and symbols, but because of its special aesthetic quality*[6], while notation is a code, related to semiotics. A graphical score is basically judged by the musicians, who comprehend, read and perform the music that lies beneath the script. In any case though, the dividing line between the two is blurry.

Concerning Logothetis's graphical scores, both terms have been used. Karkoschka for example classifies them in the "musical graphic" category. On the other hand, Logothetis himself referred to his system as a "graphical notation system", since his main focus was to broaden the musical script/code and not to provide a score with illustrative elements.

With this notation he tried to offer performers the possibility to enact music dynamically and unpredictably. That is why he developed three kinds of symbols in his notation system:

A. Pitch – symbols:

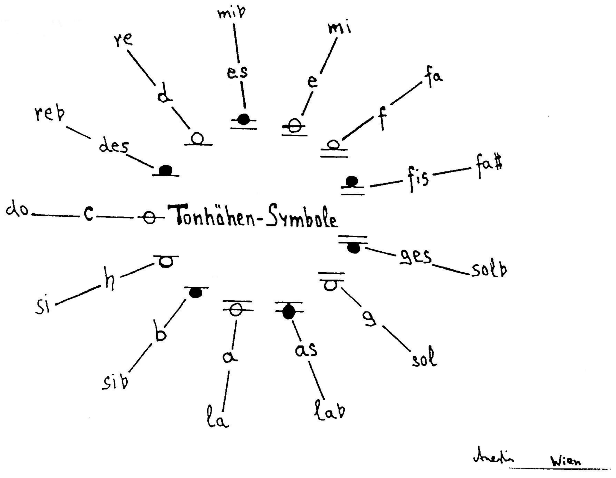


Fig. 1, Pitch - symbols

The first symbol type, seen in fig.1, consists of pitch-symbols for the realization of the tone-constellations, which can be played in every octave and combine with other markings.

B. Association – factors:

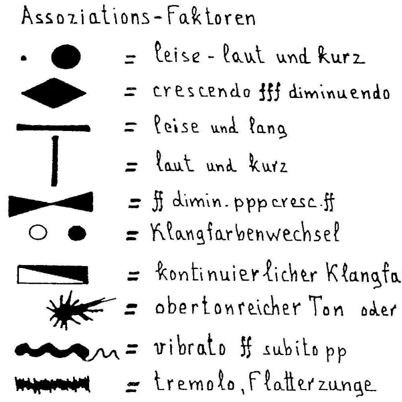


Fig. 2, Association - factors

The second category (fig.2) consists of association symbols and factors, which deal with loudness, timbre changes and sound character. The information resides in the shape of the symbol: dots mean short, and lines long note durations; loudness is derived by the size and intensity of the symbols; sound character and accent can be interpreted according to the sign's shape; the color of the signs indicates timbre changes.

C. Action signals

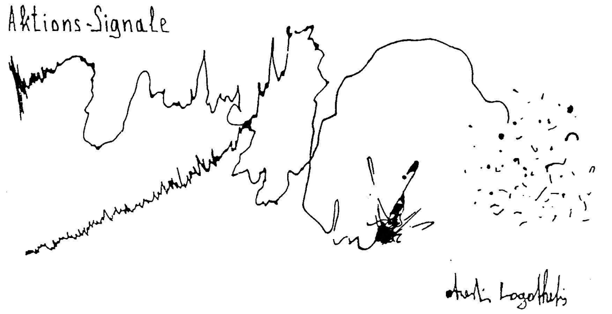


Fig. 3, Action signals

[2] Anestis Logothetis, *"Über die Darstellung des Klanges im Schriftbild", Impluse: für Spielmusikgruppen*, Universal Edition, Wien, 1973, p. 3-9

[3] Anestis Logothetis, *"Über die Darstellung des Klanges im Schriftbild", Impluse: für Spielmusikgruppen*, Universal Edition, Wien, 1973, p. 3-9

[4] Anestis Logothetis, *Zeichen als Aggregatzustand der Musik*, Wien, 1974

[5] Walter Gieseler, *"Zur Semiotik graphischer Notation"*, In: Melos NZ, Bd. 4 (1978), p. 27-33

[6] Walter Gieseler, *"Zur Semiotik graphischer Notation"*, In: Melos NZ, Bd. 4 (1978), p. 27-33

The third category (fig.3) consists of action symbols, which are lines and dots in movement. This graphical movement is to be transferred correspondingly to instruments (for transformation into music).

In the score these three kinds of symbols are combined, while the composer is trying to capture on paper his musical idea. Below there is a description example, given by Logothetis himself, of the symbols and their combinations.

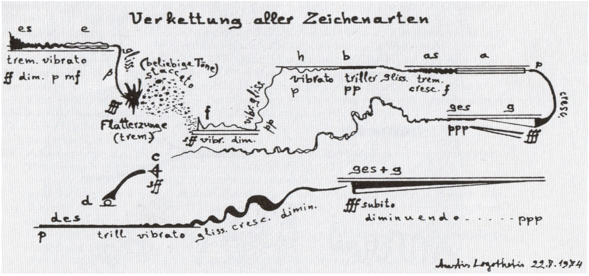


Fig.4. a typical graphical score

The above example illustrates (fig.4) that, regardless of the symbols related to the dynamics and playing technique, the ledger line symbols coincide with specific notes – the pitch being defined by the performer – while symbols lacking ledger lines relate to noises with indefinite pitch. The quality and way of playing them is left to the performer's decision, so long as the noise's auditory outcome corresponds to the visual representation adopted by the composer. *All these components will be brought into relation by means of a primarily acoustic setting and its permanent intrinsic visualization. The visual results will inform about the auditive events and their character. This kind of sound organization gives performers the freedom to participate in a multilayered hierarchy of sounds and actively shape the evolutionary polymorphy of the produced sonority*[7]. This means that the composer himself defines the boundaries of the performer's liberty as an interpreter through the way in which he draws the pattern, or graphical notation.

2. A taxonomy of Logothetis graphical scores

Accordingly he has classified his works in three different categories:

- works where the order and the temporal dimension of the signs are predefined by numbers (p.e. *Mäandros, Dynapolis, Styx, Emanation I + II)*

- works where the order and the duration of the subgroups are strictly organized, but the signs of action are left upon the improvisation of the performers(p.e. *Kentra, Ichnologia, Polymeron, Polychronon, Konvektionsströme)*

- works where the improvisation is left totally upon the performer. (*p.e.. Labyrinthos, Agglomeration, Kleine Parallaxe, Entropie*)

In a first attempt to examine Logothetis' work in chronological order, we can delineate the course through which he developed his graphical notation system.

In the table 1, which is at the end of the paper, we present some of his important works, underlining the particular innovation in his graphical language. The most crucial parameters depicted are pitch symbol, action symbol, time, instrumentation, and movement in space.

[7] Anestis Logothetis, *Zeichen als Aggregatzustand der Musik*, Wien, 1974

3. Space in the Graphic Score "Odysee" (1963)

The central feature in Logothetis' graphical notation is that sound is spatially arranged on a paper sheet surface, thus determining the final structure and flow of the piece. *The outline on a piece of paper guarantees the overview of the whole form [of the music-piece]. Every detail is built by the spontaneity of the moment and its contrapunctal processing – its polymorphism. It underlies the process of the sound-event from the graphical inflexible situation to a flexible [situation] of the sound* [8]. This helps us to see the importance space had for the composer on the sheet/paper of the score, as well as in the sound representation of the piece in physical space. He notably compared his graphical sheets/papers to architectural blueprints, which represent a large spatialization scaled down on a paper surface.

In the case of Odysee(1963) the composer treats space in a micro-scale, aiming to express the adventures of Ulysses. *Odysee* combines movement and sound through the de-codification of graphical space by the interpreters as traced in the score. Thus, the multiple adventures of Ulysses are visually 'sonified' through the personal interpretation of the instrumentalists. The main score (fig.15) which describes different psychological states related to Ulysses' adventures (island of Circe, Island of the Lotus Eaters, Cyclops, etc. is complementary to the second transparency (fig. 16) which indicates the movement of the instrumentalists in order to trace their own Odysee. The lines (vectors) which indicate the movement in fig. 9 are printed in red ink on a transparency; this transparency complements the main graphic score in order to indicate the trajectory. The musicians-dancers are divided in two or three groups. The first group traces the central trajectory -the "dromos"- and is comprised by the protagonists of the performance. From the transparency (fig.16) we can also draw information about the second and third group, and about the sonic representation of the score. Simultaneously to the first group, which follows the main path (dromos), the other two groups are on the left and on the right, producing sound masses upon the figures. Every section of the path, every vector, lasts about 2,5 min and the changes are dictated by the conductor. In every case the groups can change responsibilities and parts.

In a first attempt to "read" the graphic score of Odysee (1963) we can interpret some of the sound graphics as pictorial descriptions of Ulysses' adventures, but also as deeper psychological states which the interpreters could enter. This interpretation of the graphical score could be seen as a psychological journey to maturity, as this is described by the great Greek poet Constantinos Kavafis in his poem, *Ithaca*[9]. Through this psychodynamic interpretation of the visual graphics and their sonification, the interpreters enter an evolutionary adventure of the person situated in space and time.

For example we could refer to some of his main figures in the score as an interpretation of fig.5, which indicates the strong winds that drove Ulysses and his fellows to the island of the Cyclops. The performers may interpret this section as windy, by making the movement of the number "8" in order to express an essence of 'spaciality'. Moreover, fig.6 could be regarded as the island of the Lotus eaters, and fig.7 as the wooden spear that blinded the Cyclops Polyphemus.

We could also interpret Aeolian winds by the pattern in fig.8 which leads to the island of the Laestrygons (fig.9). The fascinating environment of Circe's island is represented by peculiar star-like patterns (fig.10). The descent in Hades is represented by the sign of the cross (fig.11) and the passage through the land of the Sirens and Scylla and Charybdis is represented in fig.12. The next stop in Ulysses' journey on the island thrinacia, is represented by fig.13 (group of figures, triangles, squares etc.). Finally, upon reaching Calypsos' island and then that of the Pheacians, Ulysses is coming to the end of his journey (fig.14). At this point, some real notes appear for the first time in Logothetis's score. To quote the composer:

In order to produce specific sound configurations, I have invented sound symbols which are "liberated" from the five line system and can be fused with other symbols, allowing for flexibility in quarters and thirds.

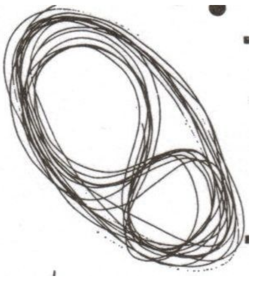


Fig. 5

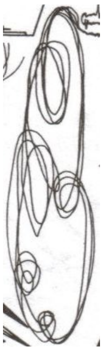


Fig. 6

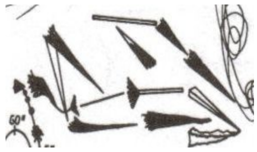


Fig. 7

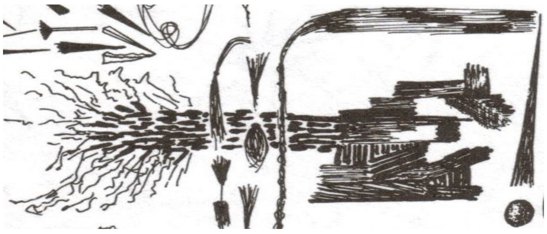


Fig. 8

[8] Anestis Logothetis, *[Zu "Polynom"]*, Text written in Zakynthos, Juni 1990, in book: Krones Hartmut, *Anestis Logothetis, Klangbild und Bildklang*, Verlag Lafite, Wien, 1998, pp. 156

[9] *"Ithaca"*

As you set out for Ithaca,
hope your road is a long one,
full of adventure, full of discovery.
Laistrygonians, Cyclops,
angry Poseidon - don't be afraid of them:
you' ll never find things like that on your way
as long as you keep your thoughts raised high,
as long as a rare excitement
stirs your spirit and your body.
Laistrygonians, Cyclops,
wild Poseidon - you won't encounter them
unless you bring them along inside your soul,
unless your soul sets them up in front of you.

http://cavafis.compupress.gr/kave_17b

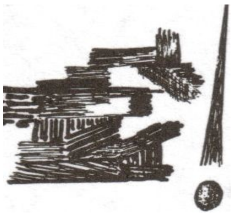


Fig. 9

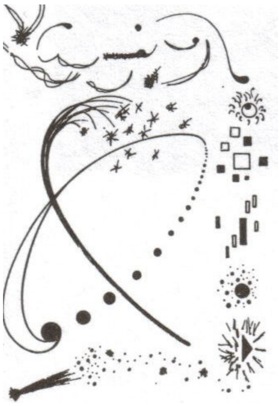


Fig. 10

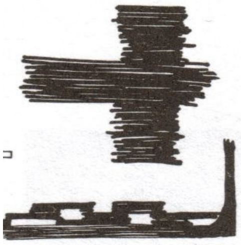


Fig. 11

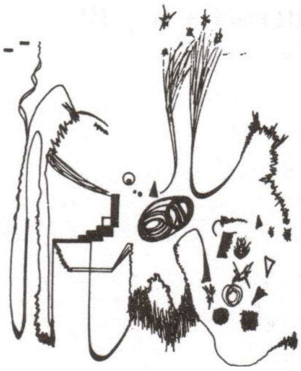


Fig. 12

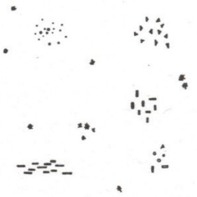


Fig 13

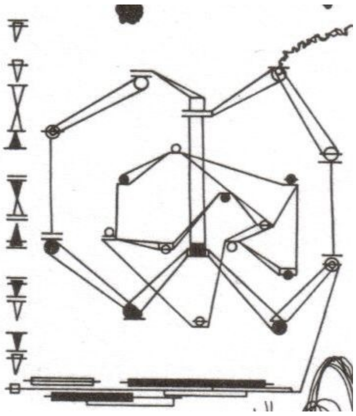


Fig.14

Conclusions

At this point our approach in the decoding of Logothetis' music, {as this is represented by the graphical notation of the composer, and its superbly dramatic enactment in the work *Odysee*(1963),} reaches a conclusion. We can state that his music and notation system form an integrated whole, that should be perceived as a unified work of art. Future research will attempt to convey and 'translate' the virtual sonification and dramatisation of the score via the means of novel tools in audio technology.

Digital technology can be used to homogenize diverse sources through the screen of the computer. All sources, visual, textual and sonic, can be translated into digital information, in order to enable digital manipulation.

References

[1] Alexaki, Eugenia, *Multimediale Tendenzen: Logothetis, Takis, Xenakis*, Dissertation FU, Berlin, 1996.

[2] Gieseler, Walter, *Zur Semiotik graphischer Notation*, Melos/NZ, Bd. 4, 1978, pp. 27-33.

[3] Goebels, Franzpeter, *Gestalt und Gestaltung musikalischer Grafik*, Melos 72/1, 1972, pp. 34.

[4] Karkoschka, Erhard, *Das Schriftbild der Neuen Musik*, Celle, Moeck Verlag, 1966, pp. 80, 128-131.

[5] Krones, Hartmut, *Anestis Logothetis, Klangbild und Bildklang*, Verlag Lafite, Wien, 1998.

[6] Logothetis, Anestis, *Kurze Musikalische Spurenkunde, Eine Darstellung des Klanges*, Melos 37/2, 1970, pp. 39-43.

[7] Logothetis, Anestis, *Über die Darstellung des Klanges im Schriftbild*, Impulse (rote reihe 34), Universal Edition, Wien, 1973.

[8] Motte-Haber, Helga de la, *Musik und bildende Kunst, von der Tonmalerei zur Klangskulptur*, Laaber, 1990.

[9] Roschitz, Karlheinz, *Anestis Logothetis und die Musik-Graphik*, protokolle – Zeitschrift für Literatur und Kunst, 1969, pp. 167-175.

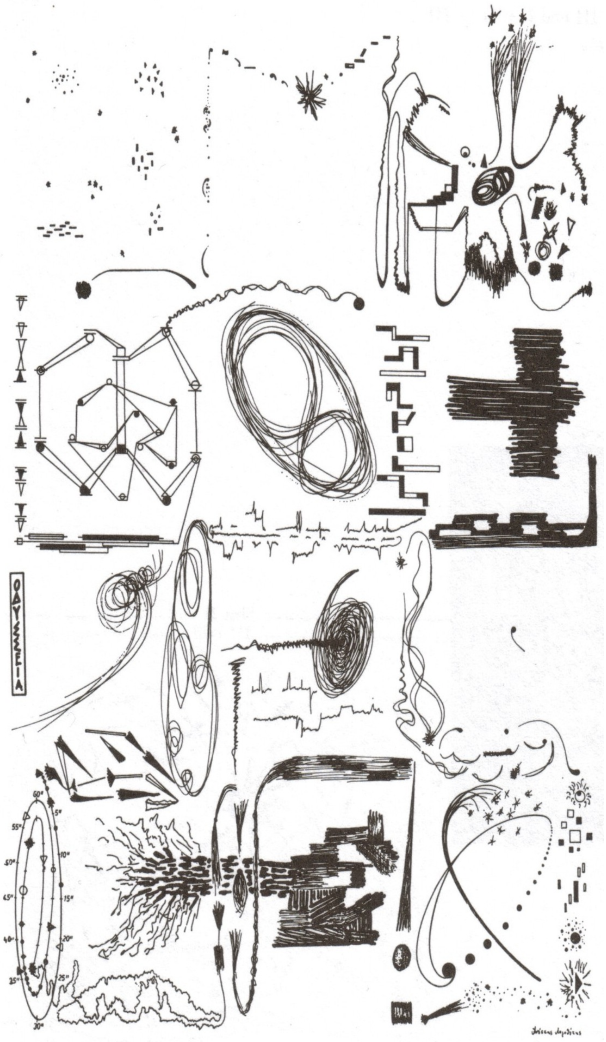


Fig. 15, The graphical score of Odysee

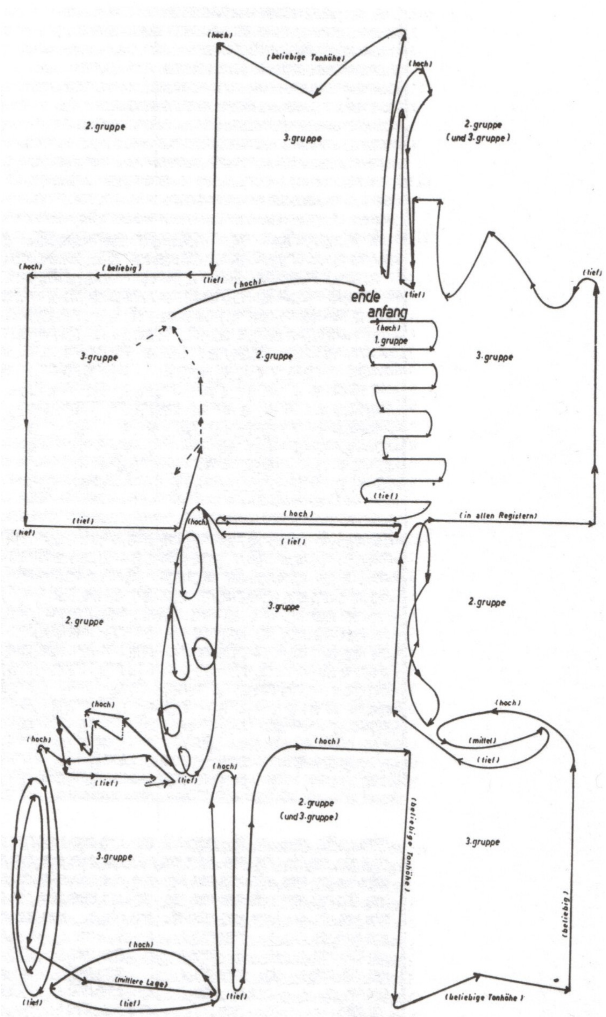


Fig. 16, The Movement of the instrumentalists (transparency)

TABLE 1

	PIECE	PITCH SYMBOLS	ACTION SYMBOLS	TIME	INSTRU-MENTATION	MOVEMENT IN THE SPACE
1957-58	**Polynom**	notes on 5lines	composed on diagramms	-	Orchestra in 5 groups	
1959	**Struktur – Textur – Spiegel – Spiel**	notes on 5lines, pitch-symbols	“D”structure gestural signal	-	variable instruments	Microstructures, no movement
1960	**Parallaxe**				Ensemble “die Reihe”	no route given
1960	**5 Porträts der Liebe: Katarakt I+II, Verkettungen I+II, Cycloide I+II+III, Novae, Reflexe I+II**	many notes, (seperate score)	many actions, (separate score)	-	Ballet (variable instruments)	route for each player, but not for the piece. -Complexity-
1960	**Agglomeration**	clear tones some chords	not many not complex	-	Solo Violin with or without Stringorchestra	given separate voice's route, but not piece's route
1961	**7 Kooptationen**	in chords	separated to pitch symbols	-	Ensemble “die Reihe”	microstructures without order
1963	**Mäandros**	12-tone piece, defined notes	in relation to pitch-symbols	Duration: 25'	Orchestra (variable instr.)	given route of the piece
1963	**Odysee**	some few tones	Many action-assosiation symbols	-	Ballet (variable instr.)	given route of the piece on a transparent paper
1963	**Dynapolis**	some defined notes	basic part of piece	Duration: 12'	Orchestra (variable instr.)	given route of piece (choice)
1964	**Ichnologia**	notes on 5lines, pitch-symbols	separate or related to pitchsymbols	-	Ensemble (variable)	given route but not in the whole work
1965	**Spiralenquintett**	undefined pitches	actions score	-	5 variable Intruments	voices' routes
1966	**Intergration**	one long chord, 12-tone	soundquality changes	-	Orchestra Groups (variable instr.)	horizontal, right + left
1961-67	**Karmadharma-drama**	undefined pitches	action-in-room score	-	Puppet choir	move in space, room-models
1968	**Styx**	12-tone piece, difened notes	in relation to pitch-symbols	Duration: 10'	Plucked Orchestra or (variable instr.)	given route of the piece
1971	**Kybernetikon**	language game – word game	in relation to words	about 30-50'	Radio Play for speaker	given reading direction
1972	**Kerbtierparty**	few notes	in relation to words	-	Radio Play	given some reading direction
1976	**Geomusik**	defined notes	in relation to pitch-symbols	Duration: 19'	Solo & Orchestra	space relativity music processes
1976-78	**Daidalia**	Pitchsymbols and text/words	Actions - Dramaturgy	-	Multimedia Spoken-Opera	
1982-84	**Aus welcem Material ist dr stein von Sisyphos**	Pitchsymbols and text/words	Actions - Dramaturgy	Given separate parts' duration	Multimedia Opera	defined movements on space
1987	**Kyklika**	12-tones in circle	quality-action changes on every tone	-	Symphony of cyclic counterpoint	given voices' route

CompScheme: A Language for Composition and Stochastic Synthesis

Luc Döbereiner
Institute of Sonology, The Hague, The Netherlands

***Abstract*—In this paper, I present a programming language for algorithmic composition and stochastic sound synthesis called *CompScheme*. The primary value generating mechanism in the program are *streams*, which allow the user to concisely describe networks of dynamic data. Secondly, I present *CompScheme*'s event model, which provides a framework for building abstract structural musical units, exemplified by showing *CompScheme*'s functionalities to control the *SuperCollider* server in real-time. Thirdly, I discuss *CompScheme*'s stochastic synthesis functionality, an attempt to generalize from I. Xenakis's *dynamic stochastic synthesis* and G.M. Koenig's *SSP*.**

I. Introduction

CompScheme is a program for algorithmic music composition and stochastic sound synthesis written in Objective Caml (*OCaml*) [6]. *CompScheme* can be used in two ways, as a library for developing applications in *OCaml*, or by accessing its functionality interactively through an interface language. All the code examples in this text are written in the interface language. The primary value generating mechanism in the program are streams, which are a concept from functional programming, which allow the user to concisely describe networks of dynamic data. Streams themselves are rules for generating values. Since streams can be combined and even used to generate new streams, the rules themselves become the object of composition. "Composing with rules" is thus not only interpreted as the mere application of a rule but the actual composition of rules, an idea that is very prominent in functional programming if one sees rules as functions.

The interface language of *CompScheme* is an implementation of the functional programming language *Scheme*[1] [1]. Instead of having a graphical environment or a fixed work flow, where the user can either visually or by filling out forms or questionnaires construct networks and derive musical data, *CompScheme* requires the user to have a degree of programming proficiency. By using an elegant, popular, small, and powerful general-purpose language such as *Scheme*, the user has all of its expressiveness and means of abstraction at his or her disposal. The user can develop full-range programs, or make small experiments by plugging together built-in streams and output functions.

Internally, *CompScheme* consists of several modules, which contain functions for specific fields of application. Data generated in *CompScheme* can be written out in several ways, such as in Wav audio files, Midi files, and binary OSC files for *SuperCollider*. It is also possible to control the *SuperCollider* server in real-time, plot and draw data. Furthermore, *CompScheme* has an event type system, which features built-in event types, and the possibility to create custom event types by bundling named parameters, setting default values, creating transformation and output functions.

CompScheme runs on *Mac OS X* and *Linux*. The top-level interpreter can run as a command line program, in a *Scheme*-mode in an editor such as *Emacs*, or on *Mac OS X* in a specially developed *Cocoa*-application, which follows the usual editor-and-listener design. A beta version of *CompScheme* can downloaded from http://sourceforge.net/projects/compscheme/.

In this text, I discuss the concept of *streams*, as well as some issues and design ideas of *CompScheme*, which relate to *streams*. Subsequently, I present *CompScheme*'s event model, which provides a framework for building abstract structural musical units. The discussion of the event model is followed by a concrete example, the modeling of the first structure of Gottfried Michael Koenig's piano piece *Übung für Klavier*. Subsequently, the expressiveness of multi-layer event streams is exemplified by presenting *CompScheme*'s facilities for controlling the *SuperCollider* server (SC Server) in real-time. Finally, the last section deals with *CompScheme*'s functions for stochastic synthesis. Starting from finding a generalization of G.M.Koenig's *SSP* and I.Xenakis's *Dynamic Stochastic Synthesis*, I have tried to develop a framework, which facilities experimentation in this field. We will see that, here, the event model is applied to a lower level, the digital sample itself.

II. Streams

As H. Abelson and J. Sussman state in their book *Structure and Interpretation of Computer Programs*, "programs must be written for people to read, and only incidentally for machines to execute."[1] Programming languages are primarily tools to express ideas. The formal nature of programming languages stipulates abstraction and generalization. Thus, through programming the structure of an idea may be revealed. Our means of expression shape what we can express. As Ludwig Wittgenstein famously formulated, "die Grenzen meiner Sprache bedeuten die Grenzen meiner Welt." ("The limits of my language mean the limits of my world.")[10] Music composition programming languages, therefore, influence our ideas

[1] The implementation of the interface language is based on *Schoca*: http://home.arcor.de/chr_bauer/schoca.html

of music. It may, thus, be argued, that the choice of a programming language also has musical consequences.

A musical performance or playback is a continuous stream of sound. In any computer representation this continuum, however, is broken up into a discrete sequence of values. The common digital representation of sound in form of a sampled waveform, common practice musical notation, as well as event-based higher-level musical abstractions follow this rule. *CompScheme* is built around the data type *streams*, which are an elegant and simple way of dealing with sequences of values, that is widely known, used, and "one of the most celebrated features of functional programming."[9] Whereas in most imperative and object-oriented systems these sequences are usually created by some iteration that collects the values or a mechanism that involves change of state, streams are persistent.

> Stream processing lets us model systems that have state without ever using assignment or mutable data. This has important implications, both theoretical and practical, because we can build models that avoid the drawbacks inherent in introducing assignment.[1]

This persistence also has advantages in musical applications. The main one, of course, is that no values are lost and everything that has been produced, and therefore everything that *will* be produced, can be referred to, which provides the user with the possibility to look into the "future" of a process and make decisions depending on what is going to happen.

Streams are flexible, they can be combined, can contain values of any kind, such as other streams or functions, and are collections as well as generative mechanisms. In more imperatively oriented systems, values are usually generated with an iterative process and collected in lists. In order to combine several processes, one has to generate values, collect them, operate on the collection, collect again, and so forth. Streams operate differently in that they generate values on demand. If several streams are combined they are piped into each other, one stream generates as much as the next one demands. Hence, infinite processes can easily be expressed. Streams, thus, allow the user to concisely describe networks of dynamic data by plugging simple parts together.

A. Finite and Infinite Streams

Most stream constructing functions in *CompScheme* return infinite streams, and it is important for the user to keep the distinction between finite and infinite streams in mind. Some operations on streams, such as plotting a stream, searching for the minimum or maximum element, accessing the last element, or appending streams require the input streams to be finite. Figure 1 shows the definition of an infinite stream and the construction of a stream which contains the first three elements of that stream appended onto the stream itself in its infinite form. The function `st-first n stream` limits the `stream` to the first `n` values, it thus converts an infinite stream into a finite one.

```
> (define my-stream1 (st-sum 1 0))
> (for-example (st-append
>                (st-first 3 my-stream1)
>                my-stream1))
(0 1 2 0 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16)
```

Fig. 1. Appending a finite and an infinite stream

B. "Inherited Ending"

When building networks of streams, in which streams act as supplies for parameter values of other streams, it is not necessary to limit the topmost stream explicitly, if a stream in the network is already finite. In other words, if any stream in a network of streams is finite, the whole network is finite and contains as many elements as the shortest stream in the network does. The stream of random values constructed in figure 2 and displayed in figure 3 ends after 10000 elements because the parameter for the lower boundary is a linear shape from 1 to 20, which ends after 10000 elements.

```
> (simple-plot
>   (st-random-value (st-line 10000 1 20) 20))
```

Fig. 2. Plotting the implicitly finite stream

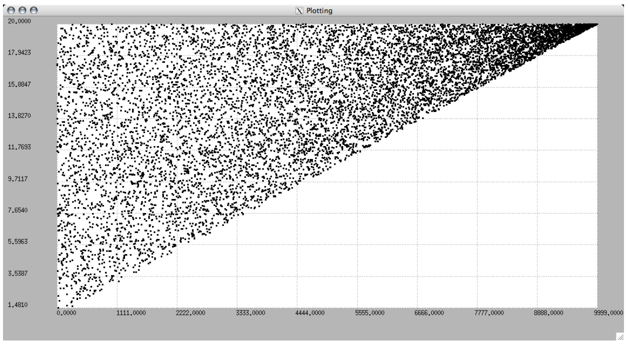


Fig. 3. The implicitly finite stream

C. Higher-Order Functions and Streams

"The most powerful techniques of functional programming are those that treat functions as data."[9, page 171] *Higher-order* functions or *functionals* are functions, which operate on other functions. In functional programming it is common to abstract by defining functions, which take other functions as arguments. This can help revealing the general structure of a method. In a stream based approach many transformations, filterings, and techniques of creating variation can be expressed in terms of higher-order functions.

Figure 4 shows the filtering of a stream of random values. The returned stream only contains those elements of the original stream, which fulfill the predicate x *mod*

12 = 0. If the numbers were to be interpreted as midi note numbers, the returned stream would only contain the pitch C in different registers. The function `st-filter function stream` returns a stream, whose elements are those of `stream` for which `function` is true.

```
> (for-example
>   (st-filter (lambda (x) (= 0 (modulo x 12)))
>              (st-rv 0 127)))
(48 120 72 96 72 120 60 48 60 72 72 108 24 24 12
24 36 72 36 0)
```

Fig. 4. Filtering a stream

Figure 5 demonstrates the function `st-apply function` $stream_1 \ldots stream_n$ by creating a triangular random distribution. The function `st-apply` constructs a stream, whose elements are the results of successively applying the function to the streams given. A triangular random distribution can be created by taking the average of two uniformly distributed random values in the same range.

```
(st-apply
  (lambda (x y) (/ (+ x y) 2))
  (st-rv 0.0 1.0)
  (st-rv 0.0 1.0)))
```

Fig. 5. Using `st-apply` to create a triangular random distribution

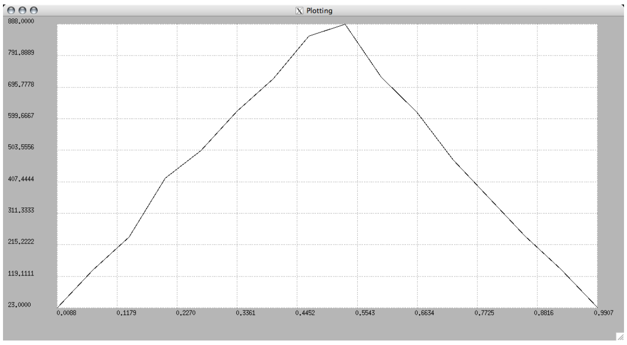


Fig. 6. A histogram of the first 10000 elements of the stream from figure 5

D. Defining Streams

Besides using the built-in streams, the user may also define his or her own streams. One way to do this is to define a function for a specific stream network. In this way, we can define a function which returns a stream of triangularly distributed random numbers by naming the stream described above, illustrated in figure 7.

```
(define (st-trirnd minium maximum)
  (st-apply (lambda (x y) (/ (+ x y) 2))
    (st-rv minimum maximum)
    (st-rv minimum maximum)))
```

Fig. 7. A simple stream definition

If the desired stream can not be built by combining already available streams, the function `st-cons value continuation` can be used. This function builds a stream, which contains `value` as its first element and `continuation` will be a delayed expression that constitutes the rest of the stream. Figure 8 shows the definition of a stream using `st-cons`. The first value is `start` and the continuation will be built by recursively calling `st-product` with the start value multiplied by `factor`, thus creating an exponentially increasing sequence.

```
(define (st-product start factor)
  (st-cons start
    (st-product (* start factor) factor)))
```

Fig. 8. A stream definition using `st-cons`

However, the definition shown in figure 8 has one limitation, being that the factor is constant and cannot be controlled using a stream. Figure 9 shows an improved version in which the factor can be dynamically controlled using a stream. In order to conform to the “inherited ending” principle, one has to check first if the factor stream is empty. For each iteration the ‘current’ value of the factor stream is accessed using the function `this` and updated for the next call using the `next` function, which returns the stream without the first (i.e. ‘current’) element. For values which are not streams, the functions `this` and `next` act as identity functions.

```
(define (st-product start factor)
  (if (empty-stream? factor)
    empty-stream
    (st-cons start
      (st-product
        (* start (this factor))
        (next factor)))))
```

Fig. 9. A stream definition using `st-cons` and a stream as argument

III. The Event Model

CompScheme’s event model provides a framework for building abstract structural units. Musical events are most commonly represented by bundling values for the description of parameters together. In this way, a musical event can be seen as a list of name-value pairs. A name may, for example, be “freq” and its associated value the number 440. *CompScheme*’s event model provides means of building name-value pairs, by defining *event types*. Event types are name-value pairs which have a name and default values. *CompScheme* has a number of general functions with which values and names can be accessed and events transformed. Events, however, do not need to be understood only as lower-level musical events, such as notes, messages to a synthesis processes etc., but may as well be representations of higher-level structural units, such as sections, passages, phrases, blocks, or entire pieces. In that sense, *CompScheme* offers a simple, yet powerful event model, which enables the free construction and aggregation of possibly inter-dependent parametric control structures on multiple temporal and structural levels.

A. Simple Event Types

Figure 10 shows the definition of a simple event type called `myevent1` with three parameters, `start`, `dur`, and `freq`, and default values associated.

```
(defevent myevent1
  (start 0.0)
  (dur 1.0)
  (freq 440.0))
```

Fig. 10. Defining a simple event type

Figure 11 demonstrates how a stream of events can be created using the function `event-stream`. In this example, the starting values are a series of numbers starting with 0.0 and increasing by 1. The frequency parameter is controlled by an exponentially distributed sequence of random numbers. The duration parameter is not controlled and will thus be the default value from the event type definition, i.e. 1.0. It is also possible for the user to provide more parameters than given in the event type definition, the event will then extend automatically and hold the additionally given values too.

```
(define eventstream1
  (event-stream 'myevent1
    (start (st-sum 1 0.0))
    (freq (st-exprand 100.0 1000.0))))
```

Fig. 11. Constructing an event stream

In general, it is the user's responsibility to define in which way an event is to be interpreted. The function shown in figure 12 can be used to output an event of the defined type in a *Csound* score file syntax with the fixed instrument number 1.

```
(define (print-myevent1 ev)
  (write "1 ")
  (map (lambda (x) (write x) (write " "))
    (list
      (event-get 'start ev)
      (event-get 'dur ev)
      (event-get 'freq ev)))
  (newline))
```

Fig. 12. Defining a printing function

B. Higher-Level Events

CompScheme also has a number of built-in event types, such as different midi and *SuperCollider* events. The functions, which output midi or *SuperCollider* events, require streams which contain events that contain at least all of the parameters, which the respective built-in types have. They may, however, contain additional name-value pairs. An event in *CompScheme* is not only to be created in the last instance, as a bundling of values before the data is written out, but may also be a representation of a higher-level structural element, which requires further interpretation. Thus, a hierarchy of events may be created and top-level or intermediate-level events and their interpretation can be created and changed independently.

The event type defined in figure 13 stands for a higher-level construct, a section. In this simple example, a section has five properties: a time offset (`offset`), a duration (`dur`), a minimum frequency (`freqmin`), a maximum frequency (`freqmax`), and starting frequency (`freqstart`).

```
(defevent mysection1
  (offset 0.0)
  (dur 10.0)
  (freqstart 100.0)
  (freqmin 100.0)
  (freqmax 10000.0))
```

Fig. 13. Defining a simple higher-level event type

However, an event is only given meaning through interpretation. Figure 14 shows the definition of a function, which constructs a stream of events of the above defined type `myevent1` with the parameters from a given event of type `mysection1`. The function `st-until` ends the "section" when the start value of the lower-level event stream is greater than the duration specified in the higher-level event.

```
(define (interpret-mysection1 section)
  (let ((offset (event-get 'offset section)))
    (st-until
      (lambda (event)
        (> (- (event-get 'start event) offset)
           (event-get 'dur section)))
      (event-stream 'myevent1
        (start (st-sum 0.1 offset))
        (freq
          (st-walk
            (event-get 'freqstart section)
            (st-rv -200.0 200.0)
            (event-get 'freqmin section)
            (event-get 'freqmax section)))))))
```

Fig. 14. Interpreting the defined event

Figure 15 shows the creation of twenty sections by creating the higher-level event stream and mapping the interpretation function, defined in figure 14, over it. Figure 16 shows the frequencies of the twenty created sections.

This simple example demonstrates the elegance and ease with which a number of sections can be produced from a higher-level description. The possibility to define abstract, higher-level structural units enables the composer to establish long-term relationships among sections, phrases, units, or blocks. It also facilitates the algorithmic organization of form concerning decisions.

```
(st-apply
    interpret-mysection1
    (st-first 20
        (event-stream 'mysection1
        (offset (st-sum 10.0 0.0))
        (frestart (st-rv 1000.0 5000.0))
        (freqmin (st-rv 100.0 8000.0))
        (freqmax (st-rv 100.0 8000.0)))))
```

Fig. 15. Controlling the higher-level event stream

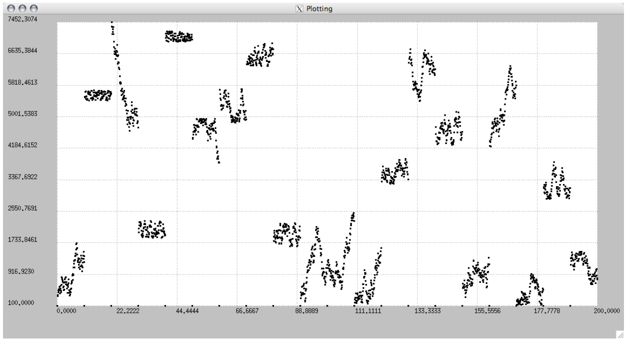


Fig. 16. The event's frequencies over their starting points

IV. Modeling Structure 1 of Koenig's *Übung für Klavier*

Gottfried Michael Koenig's piece *Übung für Klavier*, composed in 1969, is the first piece he realized with his program *Project 2*. *Project 2* was not designed for the realization of a single piece, but as a general composition program. As Koenig writes in the preface of the score, it is therefore, "necessary to generalize individual composing habits; an attempt must be made to formulate a theory – however limited – of composition."[3] *Übung für Klavier* (Study for Piano) is thus a first test of this theory. The word *Übung*, meaning 'study', but also 'practice', does not primarily refer to the player or the instrument, but rather to the composition of the piece; it is a study in writing a piece with the program *Project 2*. The title thus identifies the work as a test object and reveals the critical reflection of the work itself on the model and material from which it is derived. The piece's first section, or *structure*, will here again serve as a test object for testing the capabilities and design of *CompScheme*.

The piece consists of 12 structures and 3 variants of each of these structures, of which the pianist chooses one variant for each structure to be played. Hence, there are $3^{12} = 531441$ possible combinations of variants, of which one is performed.

Project 2 is based on a number of basic notions: compositional rules, musical quantities (*data*), characteristics of musical sound (*parameters*), combinations of rules and data (*structure formula*), and "combinatorial possibilities"[5] resulting from a structure formula (*variants*). As Koenig states, "The purpose of PROJECT 2 (PR-2) is to 'calculate musical structure variants'."[5] Since aleatoric decisions are employed in different phases of the program, it is not necessary to enter additional data for the creation of *variants* from a *structure formula*. The rules and the set of data are fixed for a specific structure and the computer constructs variants.

By following a questionnaire of over 60 questions, the *composer* describes a certain *model* of which *variations* are created. In *Übung für Klavier*, the 12 structures are *structure formulas*, variants are created by the use of aleatoric procedures. There are eight parameters that describe a *structure formula*: instrument, harmony, register, entry delay, duration, rest, dynamics, and mode of performance. Since the piece is only for one instrument, the instrument parameter is ignored.

The basic principle for the construction of musical data in *Project 2* is a three-layered process of entering, grouping, and selecting elements, the so-called *List-Table-Ensemble* principle. The construction of data for almost all parameters follows this principle. In the first instance, the composer enters a list of "allowed" elements; a basic reservoir of the smallest components. In the second layer, the user forms groups of these elements in a table, a list of selections from the list of "allowed" elements. Consequently, an *ensemble* is formed by selecting groups from the table. Thus, there are three layers of selection, i.e. choices of elements from a given supply. The first two selections are done by the composer. The composer chooses the basic elements and determines their grouping in the table. These selections and groupings remain the same for all variants of a structure. The third selection, however, is done with the help of the selection programs *alea*, *series*, and *sequence*, which chose the number and indices of the table-groups to be inserted into the *ensemble*. The third level differs from the first two levels of selection in that the selection can be changed for each variant and not single elements, but whole groups of elements are selected.

The *List-Table-Ensemble* principle is an extension of the series as a basic building block, which can be permutated in order to derive relationships. Aside from the input of the initial reservoir of "allowed" values, the other levels operate on indices. The concrete values are substituted by pointers to concrete values. The operation on pointers is an abstraction, which constitutes an intermediate meta-level, through which aspects of the musical reality become controllable. In doing so, numbers operated on never refer to themselves, i.e. calculations and the construction of numerical structures are not done for their own sake, but always for the purpose of referring to concrete values. Thus, numerical values always serve the description of musical situations. The translation of concrete values into indices creates a level on which processing is possible. The *List-Table-Ensemble* principle clearly discriminates between the material and its order. The initial input of "allowed" values is an unordered set of possible elements and the operations on indices establish orders and groupings, thereby breaking the series up into material and sequence. In *CompScheme*, I will not directly model this work flow, I will rather divide the construction into three different steps: the definition of user supplied data specific to structure 1 of *Übung für Klavier*, the definition of functions, which model the workings of *Project 2*, especially with regard to entry delay production, and thirdly plugging together the necessary stream functions and the user supplied data to define a function, which returns the structure.

In this section, I will show how the first structure of this piece can be modeled in *CompScheme*. Due to the partly incomplete description of the structure it is not possible to regenerate the exact structure, it is, however, possible to come very close to the original, using the available documentation.

The most characteristic aspect about the first structure of *Übung für Klavier* is the use of masks for the entry delays and dynamic values. Starting very dense, the entry delays gradually get larger towards the end of the structure. As an example, I will show how the entry delays are dealt with in the *CompScheme* model of this structure, the other a parameters are handled similarly. Figure 17 shows first the definition of the entry delay list and secondly the definition of two functions, which return the indices of the lower and upper boundaries of the entry delay selecting tendency mask. The parameter `x` is a position within the structure in percent of the total duration. The function `linear-shape x list` linearly interpolates the `list`, which defines line segments in the following format: n_1 $start_1$ end_1 $\ldots n_n$ $start_n$ end_n and returns the interpolated value at position `x` of the specified linear shape.

```
;; list of basic entry delay values
(define *entry-delays*
  '(0.1 0.12 0.15 0.19 0.24 0.30 0.37 0.46 0.58
      0.72 0.89 1.11 1.38 1.72))

;; functions, which return the indices of the
;; lower and upper boundaries of
;; the entry delay selecting tendency mask.
;; the parameter ‘x’ is in percent of the whole
     structure.
(define (entry-lower x)
  (round (linear-shape x '(20 0 0 27 0 1 20 0 6
      33 6 8))))

(define (entry-upper x)
  (round (linear-shape x '(20 0 6 27 4 6 20 7 10
      33 10 13))))
```

Fig. 17. The user defined masks and basic values for the entry delays

Figure 18 shows the definition of the event generating function. As seen above in figure 17, the tendency masks are relative, i.e. they do not have a fixed number of elements, but maintain their shape for different specified durations of the structure. The goal is thus to define a function, which takes the durations, i.e. the sum of all entry delays, as an argument, and returns the entry delays, while maintaining the shape of specified tendency masks. In order to do that, the entry delay selecting function needs to know its own output (the ‘current’ time). In *CompScheme*, this can be done by using `st-iterates fun arg`, which returns a streams with the following elements: arg, $(fun\ arg)$, $(fun\ (fun\ arg))$, ...

```
(define (entry-delays duration)
  (st-until (lambda (time) (> time duration))
    (st-iterates
      (lambda (time)
        (let ((percent
                (* 100 (/ time duration))))
          (+ time
             (nth *entry-delays*
               (alea (entry-lower percent)
                     (entry-upper percent))))))
    0.0)))
```

Fig. 18. The entry delay generating function

The duration and the velocity parameter are constructed similarly, but since their values depend on the position in time, i.e. on the entry delays, they do not need `st-iterates`. The generalized selection function for masks in percent is shown in figure 19.

```
(define (selecting-mask material shape-upper
    shape-lower start-times duration)
  (st-apply
    (lambda (idx) (nth material idx))
    (st-rv
      (st-apply (lambda (time)
        (shape-lower (* 100 (/ time duration))))
                    start-times)
      (st-apply (lambda (time)
        (shape-upper (* 100 (/ time duration))))
                    start-times))))
```

Fig. 19. Generalized selection with masks in percent

Figure 20 shows the definition of the final structure generating function. As stated above, one of the powerful consequences of using relative tendency masks is that the final structure can be stretched and compressed in time. Since the function accepts a duration parameter we can produce structures of any length while maintaining the same development, the same musical gesture. Many other parameters, such as duration, velocity, and register depend on the entry delays, this is why we first define a local variable `start`, which contains the entry delays. As a consequence of the above described persistence of streams, no copying of values is necessary and all streams refer to the same sequence of entry delays, despite the indeterminacy in the process of generation. The function `interval-matrix` and `st-interval-matrix` are built-in *CompScheme* functions and deal with the *Project 2* interval transition matrix pitch model, which is not to be discussed here, but described in [5] and [3].

V. Real-Time Control of the *SuperCollider* Server

The sound synthesis and music composition programming language *SuperCollider* underwent a major change in its internal design from version 2 to 3. The so-called SC Server, a powerful synthesis engine, and the *SuperCollider* language have been separated into two separate programs and now communicate with the Open Sound Control (OSC) protocol. *SuperCollider*'s system designer James McCartney writes:

```
(define (structure1 duration)
  (let ((start (entry-delays duration)))
    (st-midi-note
      (start start)
      (duration (selecting-mask *durations*
                  durations-upper
                  durations-lower
                  start duration))
      (velocity (selecting-mask *dynamics*
                  dynamics-upper
                  dynamics-lower
                  start duration))
      (note
        (st-apply
          (lambda (pc mini maxi)
            (pc-alea pc mini maxi))
          (st-interval-matrix
            (interval-matrix
              '(0 4 5 8 9 11))
            (alea 0 11) (alea 1 11))
          (st-apply
            (lambda (time)
              (nth *registers-lower*
                (registers-lower
                  (* 100 (/ time duration)))
                ))
            start)
          (st-apply
            (lambda (time)
              (nth *registers-upper*
                (registers-upper
                  (* 100 (/ time duration)))
                ))
            start))))))
```

Fig. 20. The structure generating function

> One goal of separating the synthesis engine and the language in SC Server is to make it possible to explore implementing in other languages the concepts expressed in the SuperCollider language and class library. Some other languages that I think may have interesting potential in the future for computer music are OCaml, Dylan, GOO, and also possibly Ruby[...].[8]

CompScheme's control possibilities for the SC Server are not designed to be a replacement for the *SuperCollider* language. Synth definitions (SynthDefs), recording, and routing, for example, must still be done in the *SuperCollider* language, but control and instantiation of synths can be done through *CompScheme*. *CompScheme*'s event model for controlling the SC Server is similar to the Pattern classes and the Pbind synth control (See the *SuperCollider* help files for more information on these classes) in the *SuperCollider* language, in that it allows synths to be scheduled with certain parameters. However, it differs from the Pattern classes in several ways, allowing arbitrarily deep nesting of control streams as the parameters of a synth can be updated within one event, event streams may be directly written into an OSC binary file for non-realtime rendering, and durations and entry delays are always controlled independently. Moreover, the *SuperCollider* event type (SC event) is a regular *CompScheme* event type and can be interpreted, transformed, and written out in numerous ways.

As figure 21 shows, *CompScheme*'s *SuperCollider* synth creating event type has three default parameters: the name of the SynthDef, a starting value, which is an entry delay relative to the previous event's starting time, and the duration. It is important, that the synth will free itself, at latest after the time of the duration has passed, because *CompScheme* manages the synth's IDs, in order to be able to update the synth during an event.

```
(st-sc-event
  (synth "sine1")
  (start 1.0)
  (dur 0.5)
  (<parameter4> <value4>)
  (<parameter5> <value5>)
  ...)
```

Fig. 21. The `sc_event` stream and its default values

A. Simple SC Events

Figure 22 shows a simple synth definition (synthdef), taken from [2], which is to be evaluated in the *SuperCollider* language. The parameters, which can be controlled with *CompScheme*, are the arguments of the synthdef: freq, amp, dur, attack, decay.

```
(
SynthDef("sine1",{ arg freq = 440, amp = 0.2,
  dur = 2.0, attack = 0.25, decay = 0.25;

  var   ssTime = dur * (1 - attack - decay);
  var    attackTime = dur * attack;
  var    decayTime = dur * decay;
  Out.ar(0,
          SinOsc.ar(freq, 0,amp)
                * EnvGen.kr(
                    Env.linen(attackTime,
                              ssTime,
                              decayTime,
                              1),
                    doneAction: 2)
  )
}).store;
)
```

Fig. 22. A simple synth definition (taken from [2])

Figure 23 demonstrates how to play a SC event stream in real-time. This example also demonstrates the advantage of persistent streams and the power of higher-order functions. In contrast to midi event streams, SC event streams work with relative entry delays, not absolute starting times. This decision has been made to ensure sensible time values for real-time output. In the example shown in figure 23 the starting times are made by a random choice from a list of four values. The value 0.0 stands for simultaneous events (chords). The events last until the next event starts, which is made by using the entry delays of the start parameter and dropping the first value. There is, however, one problem. Due to the chords, events which are followed by simultaneous events will have a duration of 0.0 seconds. In order to ensure that all events last at least 0.1 seconds, a clipping function is applied to the duration stream.

```
(sc-play
  (let ((entry-delays
         (st-random-choice
           '(0.0 0.1 0.15 1.7))))
    (st-sc-event
      (synth ”sine1”)
      (start entry-delays)
      (dur (st-apply (lambda (x) (max x 0.1))
             (st-drop 1 entry-delays)))
      (freq (st-exprand 100.0 4000.0)))))
```

Fig. 23. Playing a SC event stream

B. Sub-Events

As stated above, one of the powers of *CompScheme*'s SC event type system is that events, which instantiate synths, can update the synth during an event. This means that events can not only represent note-like sound events, but also control updates, within such a sound event. In general, this mechanism works by not supplying a static value or a stream of numbers, but by supplying a *stream of streams*. Every stream in this stream of streams is then seen as a development the parameter has during the respective event. However, the streams inside must be of a certain type, namely `sc_nset`. Figure 24 shows the definition of two auxiliary functions for the creation of a sub-event stream. The first function returns a stream of `st-sum` streams. The second function returns the stream of `nset`-streams we will use in the final output. The `nset` event type holds two values, `start`, which is a starting value relative to the starting point and duration of the parent event, where 0.0 denotes the starting point and 1.0 the ending of the parent event, and `value` which is the respective value used for the update of the synth's parameter. The defined function `stream-nsets` takes three arguments, which will be streams, the number of elements for each sub-event stream, the starting points and the values themselves, which are assumed to be streams of streams.

```
 ;; a stream of streams
(define (sum-streams add start)
  (st-apply st-sum add start))

;; a stream of nset streams
(define (stream-nsets st-n st-start st-value)
  (st-apply
    (lambda (nstr strt vls)
      (st-first nstr
        (st-sc-nset (start strt) (value vls))))
    st-n st-start st-value))
```

Fig. 24. Defining auxiliary functions for sub-events

Figure 25 finally shows how the an `nset`-stream can be embedded. In the example, we create a simple SC event stream, but use the above defined function for the creating a stream of `nset`-streams to control the frequency parameter. The duration of the update streams will be randomly selected between 2 and 5, the starting points of the updates are generated by streams of streams, which all start at 0.0 and increment by a constant addition of a randomly generated value for each event between 0.05 and 0.2. The frequency of each event will thus always start at 800 Hz. The defined function takes the frequency increment per sub-event as an argument, here we call the function with a constant of 100 Hz.

```
(define (sc-nset-stream1 freqadd)
  (st-sc-event
    (start 2.0)
    (dur 2.0)
    (freq (stream-nsets
            (st-rv 2 5)
            (sum-streams (st-rv 0.05 0.2) 0.0)
            (sum-streams freqadd 800.0))))))

(sc-play (sc-nset-stream1 100.0))
```

Fig. 25. Defining and playing a stream with an embedded nset stream

C. Scheduling Event Streams

It is not only possible to extend the event model to lower levels, as described in the previous section, but also to extend it to higher levels. SC event streams themselves can also be scheduled. There is another type of event called `sc_stream_event`, which contains SC event streams and starting times as relative entry delays. In the example in figure 26 a stream of SC event streams is build by mapping the SC event stream returning function `sc-nset-stream1` defined in figure 25 over a stream of random values, which will be interpreted as frequency increments for the sub-event (see previous section). The function `st-sc-stream` schedules the stream of SC event streams, the entry delays are given by the `start` argument. The function `st-sc-stream` can also take further `st-sc-stream`'s. Therefore, there is no built-in limit and scheduled event streams can be scheduled again.

```
(sc-play-stream
  (st-sc-stream
    (start (st-rv 0.0 2.0))
    (sc-stream
      (st-apply sc-nset-stream1
                (st-rv 10 100)))))
```

Fig. 26. Playing a stream of event streams

VI. Stochastic Synthesis

The idea to synthesize sound directly by using musical procedures has been employed by composers of electronic music at least since the early 1950s. Extending the compositional control down to the micro-level, and thus being able to actually compose the sound itself, has not only been part of the basic postulate of the Köln electronic music school, but has also been a general thought in many approaches to computer generated sound until today.

In the 1970s the composers Gottfried Michael Koenig, Iannis Xenakis, Herbert Brün, and others developed systems that abandoned existing acoustic models, and tried to derive sound synthesis methods directly from compositional activities. Rather than trying to compose with sounds created on the basis of given analytical models, the sound is supposed to be the result of the compositional process itself. In 1970 G.M. Koenig described his program SSP, which was not yet implemented at that time:

> As opposed to programmes based on stationary spectra or familiar types of sounds, the composer will be able to construct the waveform from amplitude and time-values. The sound will thus be the result of a compositional process, as is otherwise the structure made up of sounds. [4]

With SSP, Koenig extended the principles used in his earlier programs Project 1 and specifically Project 2 from the level of the note down to the level of the digital sample. As basic elements amplitude and time values were specified and grouped in segments, in which they were linearly interpolated. For the selection of the basic elements, aleatoric and serial principles were used. SSP may be seen as an attempt to overcome traditional ways of representation that stem from instrumental music, and substitute them with more general descriptions, such as similarity, transition, and variation that are to be applied to the macro-structure of the form as well as to the micro-structure of the sound in one process. This is derived from the axiomatic assumption, that “musical sounds may be described as a function of amplitude over time.”[4]

Iannis Xenakis’s idea of dynamic stochastic synthesis differs from Koenig’s SSP in its initial intentions. The notion of an evolutionary process is central to Xenakis’s idea of dynamic stochastic synthesis. In dynamic stochastic synthesis, breakpoints are grouped – here in cycles of a waveform – and linearly interpolated to form an integration of macro- and micro-levels of musical time. Both approaches to stochastic sound synthesis are primarily rooted in music composition, derived from compositional activities and not in the analysis of sound.

> Any theory or solution given on one level can be assigned to the solution of problems of another level. Thus the solutions in macro-composition (programmed stochastic mechanisms) can engender simpler and more powerful new perspectives in the shaping of micro-sounds. [11]

Xenakis, Koenig, and Herbert Brün were motivated by finding ways of producing sound that are idiomatic to the means of production, the computer. Instead of emulating an instrumental or electronic paradigm, the idea of the sample as the basic musical element is inherently digital. Xenakis, Koenig, and Brn used the sample as the basic musical element in a search for ”sounds that had never before existed”[11]. Instead of the novelty of sound, the strength of this non-standard approach to sound synthesis lies in its unification of the sound production and compositional processes. It is therefore really one of representation.

In the following, I present a program that is not aimed at reimplementing, but rather an attempt to generalize from Xenakis’s and Koenig’s systems for stochastic sound synthesis and thus providing the possibilities for extensions. I try to show, that the flexibility and expressiveness of streams lends itself well not only to the description of higher-level compositional processes, but as well to the lower-level sound production. Stochastic sound synthesis is an area of application in which a basic motivation of electronic music, namely *composing sound*, demands a unified representation. This unification of the sound production and the composition process requires a previous relationship between sound and control data. However, most current sound synthesis systems and computer music languages establish a strict separation of synthesis and control data. There are, therefore, hardly any platforms today, that enable experimentation in this area.

In *CompScheme*, rather than considering sound synthesis and composition as two different domains, the same mechanisms are used to describe sound as well as higher-level control. There is no separating wall between sound and control built into the system and no limit to the level of abstraction.

A. A Generalization of Stochastic Synthesis

Both SSP as well as Xenakis’s systems group amplitude and time points together, form sequences of these groups, and linearly interpolate the breakpoints. In the case of SSP, these groups – called segments – contain elements selected from initial amplitude and time lists by using Koenig’s *selection principles*. In Xenakis’s systems, these groups are cycles of one waveform, whose elements are a deviation from the previous cycle’s elements, using stochastic processes.

In *CompScheme*, the basic sound synthesis element is the sample, which contains both a time and an amplitude value. A sample is considered an event, just like any other musical event, and can be built and transformed with the same mechanism. Figure 28 shows the function `st-sample` which uses the event type syntax shown in figure 27.

```
(make-event <name>
  (<parameter1> <value>)
  (<parameter2> <value>)
  etc ...)
```

Fig. 27. Event stream creation

The example in figure 28 creates a stream of sample events from two streams, one determining the positions of the breakpoints and one that determines their amplitude. The positions in this example are taken from a list of four integers and the amplitudes are chosen randomly between -1.0 and 1.0.

```
(st-sample
  (pos (st-of-list '(0 1 2 5)))
  (value (st-random-value -1.0 1.0)))
```

Fig. 28. A sample stream

Based on SSP, we may call the sample stream of figure 28 a *segment*. A sample's time value denotes its position within the segment to which it belongs. Segments are then collected in a stream – a stream of sample streams – which can be interpolated with an interpolation function and written out into an audio file. A segment can thus be seen as cycle in a process of dynamic stochastic synthesis, or as segment in a collection from which we can select, using a *selection principle*.

B. Example 1: Dynamic Stochastic Synthesis

For a concrete example, we turn towards implementing a simple process close to Xenakis's GENDY. Generally speaking, in GENDY several breakpoints are defined and interpolated in what could be called one cycle of a waveform. The next cycle is a deviation of the previous one. Each breakpoint and time distance follows a random walk and the total length of each cycle is also controlled.

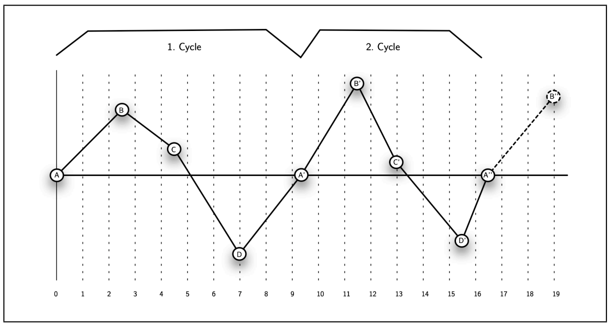


Fig. 29. Two cycles of a dynamic stochastic synthesis process

This process can easily be described in *CompScheme* using the function `segments-with-length`, which takes three arguments: the length of the cycle and two lists of numbers, the first one containing a value for each position and the latter one for each amplitude value. The time values are then scaled to fit inside of the specified length. It then returns a stream of samples; here to be considered one cycle.

```
(define (gendy1)
  (st-apply segments-with-length
    (st-walk 80.0 (st-rv -10.0 10.0) 50.0 250.0)
    (st-apply list
      (st-walk 15.0 (st-rv -2.0 2.0) 5.0 20.0)
      (st-walk 15.0 (st-rv -2.0 2.0) 5.0 20.0)
      (st-walk 15.0 (st-rv -2.0 2.0) 5.0 20.0)
      (st-walk 15.0 (st-rv -2.0 2.0) 5.0 20.0))
    (st-apply list
      (st-walk 0.0 (st-rv -0.1 0.1) -1.0 1.0)
      (st-walk 0.0 (st-rv -0.1 0.1) -1.0 1.0)
      (st-walk 0.0 (st-rv -0.1 0.1) -1.0 1.0)
      (st-walk 0.0 (st-rv -0.1 0.1) -1.0 1.0))))
```

Fig. 30. The definition of `gendy1`

Figure 30 shows the definition of a function called `gendy1`, that describes a GENDY-like process, which is kept simple for the sake of brevity. The function `segments-with-length` is successively applied to the elements of the three argument streams. The first argument controls the lengths of the cycles, by means of a random walk (`st-walk`) starting with 80, successively adding the elements of the inner stream of random values between -10 and 10 onto its current value, and limited in borders ranging from 50 to 250. The second and third arguments determine the breakpoints's positions within the cycle and are also controlled by random walks. For the sake of brevity, only four breakpoints are made. Before `segments-with-length` is applied, the time points, as well as the amplitude values, are collected in lists. It is to be mentioned, that the returned stream of waveform cycles is infinite. Since the output is a stream, we have not left the high-level description and can easily transform and reuse the created cycles.

Figure 31 shows how to write out the first 10000 cycles of a sample stream into an audio file, using a sample rate of 44100 samples per second.

```
(write-sample-stream ”gendy1.wav” 44100
  (st-first 10000 (gendy1)))
```

Fig. 31. Writing out an interpolation into an audio file

One possible extension of GENDY that composer Sergio Luque proposed [7] is the concatenation of several independent GENDYs. Similar to SSP's *permutation* function in which segments are concatenated by using *selection principles*, Luque concatenates waveform cycles from several independent GENDYs. We could easily concatenate several `gendy1s` with the function `st-interleave`, which interleaves the output of any number of streams and forms a new stream as shown in figure 32.

```
(st-interleave (gendy1) (gendy1) (gendy1))
```

Fig. 32. Concatenating three independent cycle streams

Usually stochastic synthesis is implemented in a form, that makes the positions of breakpoints depending on the sample grid. That means, that a breakpoint can only be set at a sample point. A consequence of restricting the positioning of the breakpoints to sample points is that one can only express cycles of durations, which are an integer multiple of the duration of one sample in the chosen sample rate. This limitation imposes a strong frequency grid, which is especially audible with high frequencies. A restriction like this would be considered intolerable in the case of standard oscillators, but it has been often neglected in the discussion of dynamic stochastic synthesis, in favor of reimplementing truthful adaptations of its historic original, including all of its idiosyncrasies. As can be seen in figure 30, the breakpoints's locations are expressed in floats. That means, they can be located in between samples, the resulting wave is then 'sampled' again during the interpolation process.

C. Example 2: A Variation on SSP

The following example demonstrates the use of higher-order functions to create variations of streams. In SSP, one defines segments and then selects an order of the defined segments with a function called *permutation*. In *CompScheme* there is a function called `segment` that takes three arguments: the length of the segment, a stream of relative time distances, and a stream of amplitude values and returns a stream of samples. Figure 33 shows the creation of a stream of variations of segments. The described function `segment` is mapped over three other streams, the first one producing the lengths, the second one a stream of streams produced by varying a stream, and the last argument is a stream of amplitude value streams. This last stream of streams is again produced by a mapping of a function, namely `st-repeat`, which takes two streams one containing the number of repetitions and the other containing the values to be repeated. Thus the variable `segments` contains an infinite stream of segments. Whereas in SSP every segment has to be created 'by hand', here we can easily employ SSP's principles on a higher level and create possibly infinite streams of segments.

```
(define segments
  (st-apply segment
    (st-random-value 5 50)
    (st-apply st-random-value 1 15)
    (st-apply st-repeat
      (st-apply st-random-value 1 10)
      (st-apply st-random-value -1.0 1.0))))
```

Fig. 33. The definition of the segment function

In order to create a *permutation*, we can select segments from the above defined stream by using another stream. Figure 34 shows a possible permutation. Three thousand segments are selected from the above defined `segments` with a tendency mask going from between 0 and 0 to 40 and 60 and an indexing function. Since the deviation among the first elements is smaller then that among the later ones, the output develops from a rather pitched sound to something more noisy.

```
(st-nth segments (st-tendency 3000 0 0 40 60))
```

Fig. 34. Constructing a permutation

Acknowledgment

I would like to express my gratitude to Paul Berg of the Institute of Sonology for his comments, help, and supervision, Gerhard Eckel of the Institute of Electronic Music and Acoustics in Graz for his ideas and help, and Graham Flett for his corrections.

References

[1] H. Abelson, G. Sussman and J. Sussman, "Structure and Interpretation of Computer Programs," *MIT Press*, 1996.

[2] P. Berg, "Using the AC Toolbox", *Institute of Sonology*, 2007

[3] G.M. Koenig, "Übung für Klavier", *TONOS Musikverlags GmbH*, 1969

[4] G.M. Koenig, "The use of computer programmes in creating music," *La Revue Musicale*, 1970.

[5] G.M. Koenig, "Project 2, A Programme for Muscial Composition", *Electronic Music Reports*, *Institute of Sonology*, Utrecht, 1970

[6] X. Leroy, *The Objective Caml System: Documentation and user's manual*, *INRIA*, 2007

[7] S. Luque, *Stochastic Synthesis: Origins and Extensions*, Master's thesis, *Institute of Sonology*, 2006

[8] J. McCartney, "Rethinking the computer music language: Super-Collider", *Computer Music Journal*, vol. 26, no. 4, pp. 61–68, 2002

[9] L.C. Paulson, "ML for the Working Programmer," *Cambridge University Press* , 1996.

[10] L. Wittgenstein, "Tractatus logico-philosophicus", *Suhrkamp*, 1960

[11] I.Xenakis, "Formalized Music," *Pendragon Press* , 1992.

A dynamic spatial locator ugen for CLM

Fernando Lopez-Lezcano*
nando@ccrma.stanford.edu
*CCRMA, Stanford University

Abstract— **Dlocsig is a dynamic spatial locator unit generator written for the Common Lisp Music (CLM) sound synthesis and processing language. Dlocsig was first created in 1992 as a four channel 2d dynamic locator and since then it has evolved to a full 3d system for an arbitrary number of speakers that can render moving sound objects through amplitude panning (VBAP) or Ambisonics. This paper describes the motivations for the project, its evolution over time and the details of its software implementation and user interface.**

I. Introduction

CLM is a very powerful synthesis and sound processing language in the style of Music N languages written originally in 1989 in Common Lisp by Bill Schottstaedt[2] (it was optimized at that time for running on the NeXT computer and its built in Motorola DSP). I started using it for music composition at the end of 1990, shortly after its creation.

Regretfully the NeXT workstation only had CD quality stereo outputs, which was a "downgrade" from the SamsonBox[12] four channel output, so the original version of *dlocsig* was created for the QuadBox, an external four channel D/A converter connected to the DSP port of the NeXT. The original QuadBox hardware was designed by Atau Tanaka at CCRMA and the firmware and playback software for the NeXT was programmed by myself while working at Keio University in Japan in 1992 - an "across the Pacific" joint project made possible by the Internet.

CLM[1] included "locsig", a simple panning based ugen for stereo signal location. It fell short of my needs, so I started writing another unit generator ("dlocsig", **d**ynamic **locsig**) that would encapsulate all the behavior needed to simulate most spatial cues of moving sound objects, and most importantly would be a drop in replacement for locsig so that it would be easy to modify existing instruments to use it.

Work in *dlocsig* started in 1992 and continues to this day, and the unit generator has been used by myself and other composers in numerous pieces. The original version was a four channel two dimensional system and used pair-wise panning between adjacent speakers[3]. An Ambisonics rendering back-end was added in 1999 (for B-format output and pre-rendered Ambisonics for known speaker configurations). In 2000 the ugen was extended to cover 3d arbitrary arrangements of speakers and included 3d Ambisonics and VBAP for amplitude panning. It has been part of the CLM distribution for a long time now (I don't remember when it was originally incorporated). In 2006 Bill Schottstaedt changed the name of the basic unit generator to move-sound while providing backwards compatibility synonyms within CLM.

CLM currently supports Scheme, Ruby and Forth languages in addition to the original Common Lisp, and *dlocsig* has been ported to the Scheme and Ruby worlds. The software is GPL and all source is available as part of the CLM distribution.

II. The unit generator

The current unit generator can generate spatial positioning cues for any number of speakers which can be arbitrarily arranged in 2d or 3d space. The appropriate speaker configuration is selected based on the current number of channels in the output stream. In pieces which can be recompiled from scratch this feature allows the composer to easily create several renditions of the same piece, each one optimized for a particular listening environment and rendering technique. Each user-defined speaker arrangement can also include delay compensation for the speakers and can map each speaker to an arbitrary channel in the rendered output stream.

The unit generator has several back-ends for rendering the output sound file with different techniques. The default is amplitude panning between adjacent speakers (between adjacent speakers in 2d space or three speaker triads in 3d space using VBAP[4]). It can also create an Ambisonics[5] first order b-format four channel output sound file suitable for feeding into an appropriate decoder for multiple speaker reproduction. Or it can decode the Ambisonics encoded information on-the-fly to an arbitrary number of output channels if the speaker configuration is known in advance.

An additional back end that can render 3d movements over stereo speakers or headphones using HRTFs was designed and coded in 2001 but was never finished or released.

All existing rendering back-ends can be combined while rendering a piece, and all of them take care of corner conditions like diagonal paths that cross 0,0,0 by appropriately changing the rendering details. In all cases standard cues like Doppler, multichannel output to a reverberator, amplitude scaling due to distance for the direct and reverberated components of the sound (with user-defined exponents) and ratio of direct to reverberated sound are also automatically generated.

A. Implementation

Like the rest of CLM, the original *dlocsig* core was written in Common Lisp. It is actually a Common Lisp macro that generates Lisp code on the fly to be inserted into the run loop of the instrument (the "run loop" is the section of a CLM instrument that generates the samples).

CLM unit generators are usually created and used through two functions or macros. One creates the unit generator data structures and is executed at the beginning of the instrument run, the other is usually a macro that executes the ugen code for each sample to be generated and is connected to other ugens through arbitrary lisp code.

For efficiency reasons the bulk of the complexity of the ugen was shifted to make-dlocsig, the ugen creation function. Its output is a list of parameters which the ugen itself uses to render the localized samples. The ugen itself does not know anything about rendering methods or trajectories in space and currently only knows how to apply individual amplitude envelopes to each of the output channels.

This is the Lisp definition of make-dlocsig and all its parameters (default values and some additional parameters omitted for brevity):

```
(defun make-dlocsig (start-time duration
   path
   scaler
   direct-power inside-direct-power
   reverb-power inside-reverb-power
   reverb-amount
   initial-delay
   unity-gain-dist
   inside-radius
   minimum-segment-length
   render-using)
```

start-time and *duration* define the start and duration of the sound being rendered, *path* is a path object (see below), **-power* arguments can be used to control the power exponent for attenuation due to distance for both the direct signal and the reverberated signal, and *render-using* defines the type of rendering to be done (VBAP amplitude panning, Ambisonics, etc). *inside-radius* defines the diameter of the sphere where limiting of the output signal amplitude is done and *unity-gain-dist* defines the distance at which unity gain scaling is done for the input signal.

A call to make-dlocsig returns a list that contains all the information needed by the unit generator to render the sound and the values for the start and end indexes of the enclosing run loop, as its start and duration can be affected by radial velocity and the Doppler effect (if the initial and final distances of the moving object differ).

The list contains (amongst other components) gain arrays for the direct and reverb signals with individual envelopes defined for each output channel based on the movement and the rendering method selected. Because of this internal rendering to amplitude envelopes, the same unit generator can render both amplitude panning and Ambisonics without any changes. The structure also contains an envelope for the radial velocity component of the movement so that Doppler can be generated through the use of an interpolated delay line in the unit generator.

B. Reverberation

Reverberation is not integrated into the dlocisg unit generator. It uses the standard CLM conventions for reverberation unit generators. The first versions of *dlocsig* used a tweaked four channel version of nrev, one of the most popular reverbs in the times of the Samson Box. The current version of *dlocsig* comes bundled with an n-channel version of freeverb which can reverberate n-channel inputs to n-channel outputs with a choice of local reverb percentage in the case of multichannel inputs.

C. Global Configuration

Several configuration variables can be set to globally alter the behavior of *dlocsig*.

- *dlocsig-one-turn*: the number that represents one turn, defines the angle units to be used
- *dlocsig-speed-of-sound*: defines the units of measurement for distance through the speed of sound
- *dlocsig-3d*: defines whether 2d or 3d speaker configurations are used by default
- *path-3d*: defines how paths are parsed when submitted as a plain list

In addition each of the parameters of make-dlocsig have default values based on global variables.

D. Speaker Array Configuration

Dlocsig can render soundfile output to any number of speakers when rendering to VBAP amplitude panning or to "rendered Ambisonics" (the output format for Ambisonics rendering is obviously independent of the speaker configuration).

The function *arrange-speakers* can be used to create speaker configurations for 2d or 3d setups. The speaker position is defined by their angles with respect to the listener. An additional delay can be specified for each of them in terms of time or distance, and speakers can be mapped to arbitrary output channels. Here is the definition:

```
(defun arrange-speakers (
   (speakers '())
   (groups '())
   (delays '())
   (distances '())
   (map '()))
```

speakers is a list of speaker positions in space defined using azimuth and elevation angles. Indexes from 0 in this list are used to explicitly define groups of speakers in 2D or 3D space (using the *groups* parameter - each group is a panning group of related speakers). *delays* and *distances* can be used to add delay compensation for individual speakers, and finally *map* can map any speaker to any output channel to generate custom output soundfiles that are adapted to a particular mapping of speakers in the final delivery system.

Dlocsig pre-defines a number of "reasonable" configurations for standard setups. Predefined configurations are indexed by number of output channels. A global variable (*dlocsig-3d*) is used to differentiate between flat 2D and 3D speaker arrangements. The number of output channels and the global variable are used to select a configuration at runtime and all the rendering is adjusted accordingly.

III. The Path Object

The movement of sound sources in space is described through path objects. They hold the information needed by the unit generator to move the source in space and are independent of the unit generator itself and the rendering technique used (ie: the composer uses a front end that is independent of the rendering technique used and number of output channels). Path objects can be reused and can be translated, scaled and rotated in 3d space as needed. There are several ways to describe a path. Bezier paths are described by a set of discrete points in 2d or 3d space that are joined by bezier segments through curve fitting. This description is very compact and easy to specify as a few points can describe a complex trajectory. Paths can also be specified in term of geometric entities (spirals, circles, etc). A user-defined function can also generate the points and incorporate them into a path object. In all path descriptions the velocity profile of the movement can also be specified as a function of distance.

A. Bezier Paths

This is the generic path creation function for paths defined through discrete points in space:

```
(defun make-path (path
   (3d path-3d)
   (polar nil)
   (closed nil)
   (curvature nil)
   (error 0.01)
   ;; only for open paths
   initial-direction
   final-direction)
```

The first argument, *path* is a list that specified the coordinates of the path the sound object will follow in space. Each component of the list is a list which describes a point in space and an optional relative velocity. The coordinates of each point can be specified in terms of cartesian coordinates (x, y, z) or polar coordinates (azimuth, elevation and distance - if the *polar* argument is non-nil). Paths can be open or *closed* (in the later case the initial and final points have to match). If a path is open both *initial-direction* and *final-direction* can be specified and will define direction vectors for the start and end of the movement.

If a velocity profile is not specified, the moving virtual object starts and ends at rest. The velocity profile is translated into absolute velocities for each segment of the movement by using a "constant acceleration" paradigm. Velocity is continuous at the segment boundaries and acceleration changes in a step function, being constant within each segment.

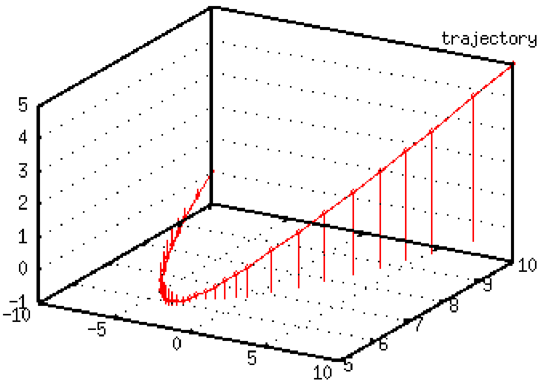


Fig. 1. Trajectory of sound object.

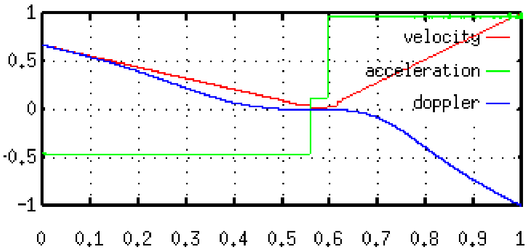


Fig. 2. Trajectory of sound object.

Here is the code that creates a very simple path expressed in cartesian coordinates:

```
(make-path
 '((-10 10 0 1)(0 5 0 0)(10 10 5 1.5)))
```

And the corresponding 3d plot of trajectory (fig. 1) and velocity, acceleration and Doppler frequency shift (fig. 2)

The path is internally rendered using a bezier curve approximation to the supplied coordinates. Each supplied point becomes a control point in the bezier curve and the control vectors are automatically derived using a curve fitting algorithm. For efficiency reasons (which may not be valid today) the rendering of the bezier curve is not done in the unit generator itself, the curves are pre-rendered to individual piece-wise linear envelopes for each output channel in the process of creating the ugen. As such, the bezier curve is approximated by individual straight line segments that are very cheap to render at sample generation time and are close enough to the original bezier curve that the Doppler shift artifacts due to the sudden change of direction at each inflection point are inaudible. The precision of the rendering process can be controlled through the *error* parameter, which defines the error bound of the linear segment approximation. The *curvature* argument controls the length of the control point vectors of the bezier curve segments so that the curvature of the bends at each control point can be controlled (see fig. 3 and fig. 4).

B. Geometric Paths

Some path subclasses exist that make the generation of some very common paths easy, in particular paths related to geometric shapes.

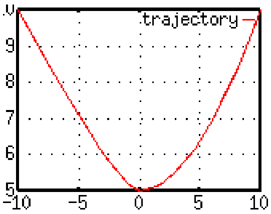


Fig. 3. :curvature '(0.4 1)

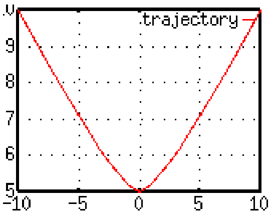


Fig. 4. :curvature 0.4

```
(defun make-spiral-path ((start-angle 0)
    total-angle
    (step-angle (/ dlocsig-one-turn 100))
    turns
    (distance '(0 10 1 10))
    (height '(0 0 1 0))
    (velocity '(0 1 1 1)))
```

Arguments should be obvious. It is possible to create arbitrary spirals from envelope like descriptions of the starting angle, total angle or number of turns and distance, with height and velocity profiles.

C. *Literal Paths*

Another class can be used to pack specific points into a path object without any further rendering or approximation done on the points in space. This makes it easy to use functions to create paths of arbitrary complexity.

D. *Path Transformations*

Path objects can be modified with some predefined transformations. They can be scaled, translated along all three axis and rotated along an arbitrary rotation point and direction vector. Paths can also be mirrored along a mirror vector. All these transformations do not affect the original coordinates of the path object, which can be reset to its original state at any time. In this way it is possible to define a set of paths and then transform them into families of paths that are used to define the movement of related sound objects in space.

E. *Path Visualization*

The path objects are created from text representations and not through a graphical editor. But methods are provided that can be used in conjunction with external visualization programs (gnuplot) to visualize the final rendition of the path as defined in the object.

It is possible to plot the trajectory, velocity profile, acceleration profile and Doppler shift of the moving sound object.

F. *Examples*

A very simple Lisp instrument that uses dlocsig:

```
(definstrument sinewave (start-time duration freq amp
    &key
    (amp-env '(0 1 1 1))
    (path (make-path
            :path '(-10 10 0 5 10 10))))
  (multiple-value-bind (dloc beg end)
      (make-dlocsig :start-time start-time
                    :duration duration
                    :path path)
    (let* ((osc (make-oscil :frequency freq))
           (aenv (make-env :envelope amp-env
                           :scaler amp)))
      (run
       (loop for i from beg below end do
         (dlocsig dloc i (* (env aenv)(oscil osc))))))))
```

This snippet of code will render one note created with the previously defined instrument in a four channel two dimensional setup:

```
(with-sound(:channels 4)
  (sinetest 0 1 440 0.5 :path
      (make-path '((-10 10)(0 5)(10 10)))))
```

The same instrument could render a 3d path in a cube of 8 speakers:

```
;; tell the system I want to use 3d
(setf dlocsig-3d t)
;; render the sound with a 3d path
(with-sound(:channels 8)
  (sinetest 0 1 440 0.5 :path
      (make-path '((-10 10 0)(0 5 10)(10 10 5)))))
```

IV. Limitations

The use of an interpolated delay line to render the Doppler frequency shift imposes a limitation to CLM instruments that use it. CLM is by nature a "sound painting" environment. It does not require notes to be time ordered in its score (which is just a Lisp program), not does it require samples to be output in time order from within an instrument (ie: any instrument can sprinkle sounds at arbitrary times in the output sound file). That absolute freedom in the time ordering domain does not mesh with a delay line that has to be fed with a constant stream of samples, so some instruments are not compatible with *dlocsig*.

An example is grani, a general purpose granular synthesis instrument I started writing in 1996. Generated grains are not necessarily time ordered and thus cannot be fed to *dlocsig* inside the instrument itself.

In these cases it is relatively easy to work around the problem by using *sound-let* and a very simple instrument that can move an arbitrary sound file in space (it is provided as an example in the CLM distribution). Sound-let calls the troublesome instrument and creates an intermediate sound file which is later spatialized by *dlocsig*. The process is transparent to the composer and the additional time overhead of the intermediate sound file creation is not significant.

V. Future Directions

There are many things in my list of "things to do" for Dlocsig, here are some details about the most important of them:

- The Ambisonics encoding back end is being expanded to include second (and higher) order Ambisonics encoding[6] [7].
- The Ambisonics rendering back end (used when the selected rendering type is *decoded-ambisonics*) is too simple, it should be extended to do dual band decoding that properly tries to match velocity and energy vectors in the low and high frequency bands. Or maybe the internal renderer should be scrapped altogether, it was merely created as a convenience and an external decoder[9] could be used in most if not all cases (for example Ambdec includes hand tuned configurations for 5.1 Ambisonics rendering[8]).
- The unfinished HRTF based back end should be finished and included in the distribution.
- It would also be interesting to explore the possibility of adding a Wave Field Synthesis back end. This would be difficult as it would imply a separate soundfile for each sound object, an approach that is at odds with the current "piece as a soundfile" Dlocsig / CLM system. *Path* objects, on the other hand, could easily generate the information to later do WFS rendering of the soundfiles.
- The Bezier curve fitting and rendering system for the *path* objects should be reconsidered to see if using a different type of curve fitting algorithm might produce better results. Bezier segment fitting can sometimes result in pathological behavior with some paths, specially with loops being created automatically. A more generic approach could use NURBs (Non-uniform rational B-splines) but fitting algorithms would have to be found.
- The tessellation algorithm described in Pulkki's VBAP paper[4] should be implemented so that grouping of speakers is automatically done.

ACKNOWLEDGMENT

Thanks to Bill Schottstaedt for creating, maintaining and expanding the wonderful CLM environment.

REFERENCES

[1] Common Lisp Music (CLM): http://ccrma.stanford.edu/software/clm/

[2] Bill Schottstaedt, "CLM: Music V Meets Common Lisp," *Computer Music Journal* 18(2):30-37, 1994.

[3] John. Chowning, "The simulation of moving sound sources," *Journal of the Audio Engineering Society*, vol. 19, no. 1, pp. 26, 1971.

[4] V. Pulkki, "Virtual sound source positioning using vector base amplitude panning", *Journal of the Audio Engineering Society*, 45(6) pp. 456-466, June 1997.

[5] Michael A. Gerzon, "Periphony: With-Height Sound Reproduction", *Journal of the Audio Engineering Society*, 1973, 21(1):210

[6] Dave Malham, "http://www.york.ac.uk/inst/mustech/3d_audio/higher_order_ambisonics.pdf", 2003

[7] Jerome Daniel, "Reprsentation de champs acoustiques, application la transmission et la reproduction de scnes sonores complexes dans un contexte multimdia", Thse de doctorat de lUniversit Paris 6, 2001

[8] Bruce Wiggins, "An Investigation into the Real-time Manipulation and Control of Three-dimensional Sound Fields", University of Derby Doctoral Thesis, 2004

[9] Fons Adriansen, Ambdec, an open source Ambisonics decoder, "http://www.kokkinizita.net/linuxaudio/downloads/ambdec-manual.pdf"

[10] Fernando Lopez-Lezcano, "A Four Channel Dynamic Sound Location System", *The Japan Music and Computer Science Society (JMACS) 1992 Summer Symposium*, 1992

[11] Fernando Lopez-Lezcano, "A dynamic spatial sound movement kit", *International Computer Music Conference (ICMC)*, 1994

[12] Julius Smith, "http://www-ccrma.stanford.edu/ jos/kna/Experiences_Samson_Box.html"

Zsa.Descriptors:
a library for real-time descriptors analysis

Mikhail Malt [*], Emmanuel Jourdan[†]
[*] IRCAM, Paris, France, Mikhail.Malt@ircam.fr
[†] IRCAM, Paris, France, Emmanuel.Jourdan@ircam.fr

I. Introduction

In the past few years, several strategies to characterize sound signals have been suggested. The main objective of these strategies was to describe the sound [1]. However, it was only with the creation of a new standard format for indexing and transferring audio MPEG 7 data that the desire to define audio data semantic content descriptors came about [2, p.52]. The widely known document written by Geoffroy Peeters [1] is an example where, even if the goal announced is not to carry out a systematic taxonomy on all the functions intended to describe sound, it does in fact systematize the presentation of various descriptors.

A. Descriptors Today

A large percentage of the uses for descriptors concern primarily indexing and browsing contents of sound databases or for re-synthesis of sounds, such as in telephone transmissions.

Our interest concerns the use of the analysis of descriptors in real-time for the creation and analysis of contemporary music. In this domain, with the exception of the fundamental frequency and the energy of the sound signal, the use of spectral descriptors is still rare. It is important nonetheless, to examine the experiments on computer-assisted improvisation carried out by Assayag, Bloch, and Chemillier [3] [4] and the developments in "concatenative synthesis" by Diemo Schwarz [5] where the analysis of a variety of sound descriptors is used to control re-synthesis.

B. Descriptors and Music Composition

The fact that descriptors are rarely used in contemporary music compositions is due to several factors including:

- The lack of knowledge of the relationships between descriptors and the pertinent perceptual characteristics of the sound for use in musical composition;
- The fact that one descriptor is not sufficient in order to characterize a complex "sound state" such as that of a note played "live." Recent studies [6] show how the composed functions of descriptors are more effective in recognizing the characteristics of a given sound signal than the use of one descriptor at a time;
- The lack of a large choice of descriptors in real-time so that artists can test them and learn to use them.

II. Real-Time Environments and Descriptors

Among the most widely used software environments for real-time musical performances are *SuperCollider* [7], *PureData* [8], and *Max/MSP* [9]. Max/MSP offers the largest selection of tools to work with sound descriptors. Currently, several libraries offering analyses of descriptors are available in Max/MSP. The best known of these environments include the library by Tristan Jehan [10] [11] (pitch~, loudness~, brightness~ , noisiness~, bark~, analyzer~, shifter~, segment~, beat~), the iana~ object of Todor Todoroff, the yin~ object implemented by Norbert Schnell, according to the Cheveigné and Kawara model [12], the *FTM/Gabor* object library [13] [14] that enables development of descriptors, and finally the classic fiddle~ and bonk~ by Miller Puckette [15].

However, a large number of the descriptors offered are, as we have already mentioned, based on the recognition of the fundamental frequency and the energy. The only exceptions are the descriptors offered in the *Gabor* library, but they do not cover yet a large set of descriptors.

III. The First Descriptors Set Available in Zsa.Descriptors

The Zsa.Descriptors library is intended to provide a set of audio descriptors specially designed to be used in real-time. This objects collection encloses a sound descriptors set coming from the MPEG-7 Descriptors, outlined by Peeters [1], algorithms for peak search from Serra [16] and some ideas from the Computer Assisted Composition developments realized by the Musical Representation Team at Ircam. In the next paragraphs we will describe some of this tools, already developed in the Zsa.Descriptors library.

A. Spectral Centroid (brightness)

This is a very well known descriptor. The Spectral centroid is the barycentre of spectra, computed as follow:

$$\mu = \frac{\sum_{i=0}^{n-1} f[i]a[i]}{\sum_{i=0}^{n-1} a[i]}$$

Where:

n, is the half of the fft window size

i, the bin index

$a[i]$, is the amplitude of the bin i, the real part of the FFT calculus

$f[i]$, is the frequency of the bin i. where

$$f[i] = i * \frac{sample\ \ rate}{fft\ \ window\ \ size}$$

and

μ, is the spectral centroid in hertz.

B. Spectral Spread (spectral centroid variance)

As usual, we consider the spectral centroid as the first moment of spectra, considered as a frequency distribution, which is related with the weighted frequency mean value. The spectral spread is the second moment, i.e., the variance of the mean calculated above.

$$v = \frac{\sum_{i=0}^{n-1} (f[i]-\mu)^2 a[i]}{\sum_{i=0}^{n-1} a[i]}$$

C. Spectral Slope

The spectral slope is an estimation of the amount of spectral magnitude decreasing, computed by a linear regression on the magnitude spectra.

$$slope = \frac{1}{\sum_{i=0}^{n-1} a[i]} \frac{n\sum_{i=0}^{n-1} f[i]a[i] - \sum_{i=0}^{n-1} f[i] \sum_{i=0}^{n-1} a[i]}{n\sum_{i=0}^{n-1} f^2[i] - \left(\sum_{i=0}^{n-1} f[i]\right)^2}$$

D. Spectral Decrease

The spectral decrease meaning is similar to spectral slope, representing the amount of spectral magnitude decreasing. According to Peteers [1], this formulation comes from perceptual studies and it is supposed to be more correlated to human perception.

$$decrease = \frac{\sum_{i=0}^{n-1} a[i] - a[1]}{\sum_{i=2:K}^{n-1} a[i](i-1)}$$

E. Spectral Roll-Off

The spectral roll-off point is the frequency $f_c[i]$ so that x% of the signal falls below this frequency. "x" is took as 0.95 as default value. The roll-off point is calculated as follow:

$$\sum_{i=0}^{f_c[i]} a^2[f[i]] = x\sum_{i=0}^{n-1} a^2[f[i]]$$

where:

$f_c[i]$, is the roll-off point and

x, the roll_off energy percent accumulated.

F. Sinusoidal model based on peaks detection

We have based the calculus of our algorithm on the widely known method defined by Smith&Serra [16] [18, p. 38-48], where a peak is defined as a local maximum in the real magnitude spectrum $a_k[i]$."k" is the frame index. As not all the peaks are equally important in the spectrum, we have used a sliding five points window to scan the magnitude spectra, avoiding undesired peaks. For each 5 magnitudes vector we check for the third point $a_k[2]$, and for a given threshold value ε_t, we compute:

$a_k[2] = \max\{a_k[0], a_k[1], ..., a_k[4]\} \wedge a_k[2] > \varepsilon_t$.

If the condition is true, then $a_k[2]$ becomes a peak.

A parabolic interpolation is then applied on the three adjacent points, $a_k[1], a_k[2], a_k[3]$.

Solving the parabola peak location [18, p. 47], a coefficient "p" of the "j" peak is then calculated:

$$p_j = \frac{1}{2}\frac{a_k[1]-a_k[3]}{a_k[1]-2a_k[2]+a_k[3]}$$

The true peak location (in bins) is given by:

$$i_{peak[j]} \equiv i_{a_k[2]} + p_j$$

To estimate the true magnitude we use $\boldsymbol{p}$ as follow:

$$a_{k\,peak[j]} \equiv a_k[2] - \frac{1}{4}(a_k[3]-a_k[1])p_j$$

At the end of the process we have collected a set of partials $pk_j = (f_j, a_j)$.

G. A Tempered Virtual fundamental

This descriptor was based on the harmonic histogram technic described by Jean Laroche [20, p.52-53]. We adapted this method in order to approximate the result and the research phase for the "best candidate", by a tempered musical scale with a given division.

Given a set of peaks $pk_j=(f_j,a_j)$, calculated as showed previously,

1) For each $pk_j=(f_j,a_j)$ we calculate a set of

$$pk_{jn}=\left(\frac{f_j}{n},a_j\right),n\in N,n\subset[1,..,6].$$

2) All the $\frac{f_j}{n}$ were converted in indexes, i_{jn}, in a pitch-class space. At this level $i_{jn}\in R$.

3) The i_{jn}were approximated by a grid of discrete values multiples of q, $q\in R,q\subset[0,1]$, returning a new set of values i^q_{jn}, multiples of q. Notice that q can be seen as a half-tone division. $q=1$, means an approximation by a half-tone, $q=0.5$ an approximation by a quarter-tone, and so on.

4) This leads us to new couples $pk^q_{jn}=(i^q_{jn},a_j)$.

5) Collecting all couples according with the identical i^q_{jn}, we build new couples $pk^q_{jn}=\left(i^q_{jn},\sum a_j\right)$, where $\sum a_j$ is the sum of all a_j for the identical i^q_{jn}.

6) The best candidate to be our virtual fundamental will be the $pk^q_{jn}=\left(i^q_{jn},\sum a_j\right)$, that maximises $\sum a_j$.

7) In the last phase, $i^q_{jn}$is converted, in floating point MIDI pitches or in a frequency space.

IV. The Solution offered by Zsa.Descriptors

As was exposed previously, the main goal of Zsa.Descriptors, a library of sound descriptors and spectral analysis tools, is to expand the capabilities of sound description using the systematic approach of the MPEG7 standard, and to offer a set of truly integrated objects for the *Max/MSP* [9] graphical programming environment. In addition to sound descriptors and original analysis features, the external objects of the Zsa.Decriptors library are designed to compute multiple descriptors in real-time with both efficiency in terms of CPU usage and guaranteed synchronization. In consequence a modular approach was chosen. In the *MAX/MSP* environment context, this was made possible by sharing the expensive process of the windowed FFT.

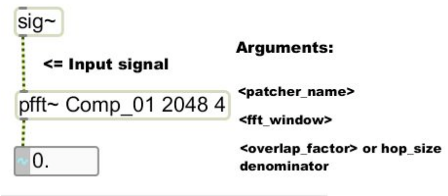


Fig. 1 the pfft~ object

MAX/MSP, actually has an object that calculates, in an efficient form, a windowed FFT, the "pfft~" facility.

As is said, in the documentation, "pfft~", is a "Spectral processing manager for patchers", i.e., an object that can load special designed "patchers". The "pfft~" object takes at least three arguments: the patcher name, the FFT window size and an overlap factor, to calculate the hop size (Fig. 1). The loaded "patcher" must also follow a general structure. This patcher must have at least an "fftin~" object. The pfft~ object manages the windowing and overlap of the incoming signal, fftin~ applies the window function (envelope) and performs the FFT. The fftin~ object takes two arguments, the "inlet assignment" and the name of the window envelop function (hanning, hamming, square and blackman are included), or the name of buffer~. It is therefore possible to use any kind of window depending on the type of sound that we want to analyse

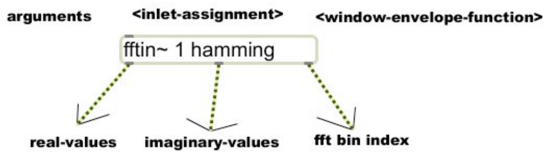


Fig. 2 the fftin~ object

Therefore, most of the objects of the library, was designed to run inside the standard *MAX/MSP* pfft~ object (Fig. 3). This strategy offers multiple advantages: modularity, efficiency, and also the ability of using the analysis directly as parameter for sound processing in the spectral domain.

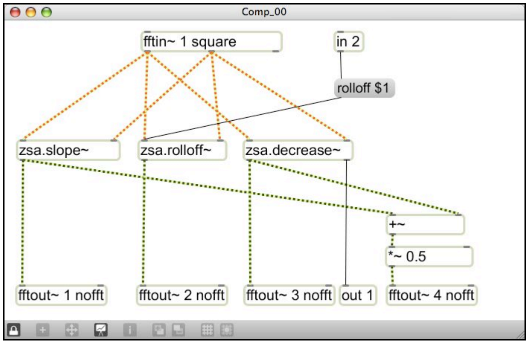


Fig. 3 Interior of the pfft~ object

Furthermore, the fact that the objects of this library can work within the Max/MSP environment either together or by themselves and the fact that they work smoothly in conjunction with other standard Max/MSP objects, makes it possible to exploit all the synchronization resources available in this environment.

V. Conclusions and Perspectives

We have presented in this paper a set of sound descriptors and signal analysis tools intended to be used in real-time as a toolbox for composers, musicologist and researchers. This will allow the use and the research on the use of sound descriptors in the fields of systematic musicology and as a tool for taking decisions in the real-time performance context. As part of the future work, we have also planed a research on the musical segmentation based on sound descriptors as a strategy to musical analysis.

The main advantage, of this library, more than the small improvements we did in some algorithms, was the modular development technique and implementation we have used, trying to optimise the calculus by a strong integration in the MAX/MSP environment.

Of course, the work presented here still preliminary, but it will be improved with the implementation of the following list of features, which are already implemented or currently being developed: temporal variation of the spectrum, bark, inharmonicity, harmonic spectral deviation, odd to Even harmonic energy ratio, tristimulus, frame energy, harmonic part energy (this harmonic descriptors will use a monophonic F0 algorithm developed by Chunghsin YEH in his Ph.D. thesis [21]), noise part energy, and others descriptors coming from the signal processing and computer assisted composition worlds.

ACKNOWLEDGMENT

We would like to thank Richard Dudas for his fruitful remarks, comments and suggestions, Arshia Cont for his friendly support and Cyril Beros for funding the travel support.

REFERENCES

[1] G. Peeters, A large set of audio features for sound description (similarity and classification) in the CUIDADO project. Cuidado projet report, Institut de Recherche et de Coordination Acoustique Musique (IRCAM), 2004.

[2] F. Gouyon, Extraction automatique de descripteurs rythmiques dans des extraits de musiques populaires polyphoniques, mémoire de DEA ATIAM, Université de la Méditérranée, Université Paris VI, IRCAM, Télécom-Paris, université du Maine, Ecole Normale Supérieure, ACROE-IMAG, Juillet 2000.

[3] G. Assayag, G. Bloch, M. Chemillier, A. Cont, S. Dubnov – « Omax Brothers : A Dynamic Topology of Agents for Improvization Learning », in Workshop on Audio and Music Computing for Multimedia, ACM Multimedia 2006, Santa Barbara, USA, October 2006.

[4] G. Assayag, G. Bloch, M. Chemillier – « OMax-Ofon », in Sound and Music Computing (SMC) 2006, Marseille, France, Mai 2006

[5] D. Schwarz, « Current research in concatenative sound synthesis », Proceedings of the International Computer Music Conference (ICMC), Barcelona, Spain, September 5-9, 2005.

[6] Zils A., Extraction de descripteurs musicaux: une approche évolutionniste, Thèse de Doctorat de l'Université Paris 6, Septembre 2004.

[7] SuperCollider, © James McCartney, http://www.audiosynth.com/

[8] PureData, © Miller Puckette, http://crca.ucsd.edu/~msp/Pd_documentation/

[9] Max/MSP, © Cycling74, www.cycling74.com

[10] T. Jehan, B. Schoner, « An Audio-Driven, Spectral Analysis-Based, Perceptual Synthesis Engine », in Audio Engineering Society, Proceedings of the 110th Convention, Amsterdam, The Netherlands, 2001.

[11] T. Jehan, Creating Music by Listening, PhD Thesis in Media Arts and Sciences, Massachusetts Institute of Technology, September 2005.

[12] A. De Cheveigné, H. Kawahara, « YIN, a fundamental frequency estimator for speech and music », J. Acoust. Soc. Am. 111, 1917-1930, 2002.

[13] N.Schnell et al. « FTM Complex Data Structures for Max/MSP », in ICMC 2005, Barcelona, Spain, 2005..

[14] N. Schnell, D. Schwarz, “Gabor, multi-representation real-time analysis/synthesis”, Proc. of the 8th Int. Conference on Digital Audio Effects (DAFx’05), Madrid, Spain, September 20-22, 2005

[15] M. Puckette, T. Apel,. « Real-time audio analysis tools for Pd and MSP ». Proceedings, International Computer Music Conference. San Francisco: International Computer Music Association, 1998, pp. 109-112.

[16] J.O. Smith, X. Serra, “PARSHL: an analysis/synthesis program for non- harmonic sounds based on a sinusoidal representation”. Proc. 1987 Int. Computer Music Conf. (ICMC’87), Urbana, Illinois, August 1987, pp. 290 -297.

[17] B. Doval, X. Rodet, “Fundamental frequency estimation and tracking using maximum likelihood harmonic matching and HMMs.” Proceedings of the ICASSP ‘93, 1993, pp. 221- 224.

[18] X. Serra, A system for sound analysis/transformation/synthesis based on a deterministic plus stochastic decomposition. Philosophy Dissertation, Stanford University, Oct. 1989

[19] X. Rodet, “Musical Sound Signal Analysis/Synthesis: Sinusoidal+Residual and Elementary Waveform Models”, in TFTS'97 (IEEE Time-Frequency and Time-Scale Workshop 97), Coventry, Grande Bretagne, august 1997.

[20] J. Laroche, Traitement des signaux audio-fréquences, TELECOM, Handout, Paris, France, February 1995.

[21] C. YEH, Multiple fundamental frequency estimation of polyphonic recordings, Ph.D. thesis, Université Paris 6, 2008.

Ephemeron:
Control over Self-Organised Music

Phivos-Angelos KOLLIAS[*]
[*] CICM, University of Paris VIII, FRANCE, soklamon@yahoo.gr

Abstract — **The present paper discusses an alternative approach to electroacoustic composition based on principles of the interdisciplinary scientific field of Systemics. In this approach, the setting of the electronic device is prepared in such a way to be able to organise its own function, according to the conditions of the sonic environment. We discuss the approaches of Xenakis and of Di Scipio in relation to Systemics, demonstrating the applications in their compositional models. In my critique on Di Scipio's approach, I argue that the composer is giving away a major part of his control over the work and therefore the notion of macro-structural form is abandoned. Based on my work *Ephemeron,* I show that it is possible to conduct emerging situations applying the systemic principle of 'equifinality'. Moreover, I argue that it is possible to acquire control over these situations and their properties over time so as to develop formal structure.**

PREFACE

I do not believe that any treatise of music aesthetics, using the rhetorical skills in the domain of language, and supported by suitable logical arguments, can suggest an absolute manner of creation and of perception or that it can promise to be more effective than others. Nevertheless, a music treatise can demonstrate the framework in which a work has come into existence and in which it can be appreciated in a clearer fashion. It can help in the work's appreciation both in the logical domain and in the purely musical domain. In respect to that, my conviction is that a study including criticism on other approaches, serves only to show the similarities and the differences between the composer's aesthetical and methodological position is willing to propose and that of other aesthetical positions. Thus, I see no interest in a polemic treatise of aesthetics other than the pleasure of polemics itself. In this sense, the current paper that includes criticism mainly on the conceptual positions of other composers, serves to connect and distinguish the approach in my work *Ephemeron* in connection to theirs.

I.INTRODUCTION

The paper discusses an alternative approach to electroacoustic composition based on principles of the interdisciplinary scientific field of Systemics. In this approach, the setting of the electronic device is prepared in such a way to be able to organise its own function, according to the conditions of the sonic environment. In this way, the music result has a unique character in each performance.

The discussion, placed in the context of Systemics, starts with an introduction of some fundamental systemic concepts. By referring to Xenakis' 'Markovian Stochastic Music', I present one of the first attempts to apply in music concepts deriving from the theory of Cybernetics. Di Scipio critique on Xenakis is examined, as it is one of the fundamental factors for the conception of his own musical application of Systemics. Di Scipio's model of 'Audible Ecosystemics' is demonstrate, in which the central role has the concept of a self-organised system. In my critique on Di Scipio's approach, I argue that the composer is giving away a major part of his control over the work by choosing to opperate only on the basic organisational level. In this sense, even though the composer controls the communication between the system and its environment, he loses control over the final result. Consequently, the notion of macro-structural form is abandoned. Moreover, I attempt to define 'self-organized music' and to establish a general model in the context of electroacoustic music. For this, I am using Di Scipio's model, interpreting it through the model of Second-Order Cybernetics.

The last section is devoted on the presentation of my compositional approach using systemic principles, through my work *Ephemeron*. I demonstrate the structure of the model based on the concept of an 'adaptive living organism' and its first complete application in the concert hall of ZKM. Through that, I show that it is possible to conduct emerging situations, applying the systemic principle of 'equifinality'. Moreover, I claim that it is possible to acquire control over these situations and their properties over time so as to develop formal structure.

II.SYSTEMICS AND MUSIC

My research is focused into Systemics for a period of more than three years. First, the interest started from a philosophical viewpoint, fascinated from the idea that everything can be considered and be observed as an organisation. Then, I focused in applying the model and its concepts into music.

As I have shown in previous studies [1] [2], Systemics can be applied in all musical creation. However, here I will limit the discussion only in self-organised electroacoustic music and mainly in connection with Xenakis' approach and more particular with that of Di Scipio.

A. Introduction to some systemic concepts[1]

Before starting the discussion around self-organised music, I will suggest some concepts of the original field of Systemics. First of all, Systemics is consisted of a number of interconnected interdisciplinary theories, mainly Cybernetics, General Systems Theory and the more recent Complexity Science. The main framework of Systemics is the treatment of organised entities. In this viewpoint, everything can be considering as a system.

In its abstract sense, as Bertalanffy explains, '*a system is a whole consisted of interacting parts'* [3]. From the perspective of *system differentiation theory*, as Luhmann explains, the division between *whole and parts* becomes *system and environment* [4]. In this sense, a part of a system can be considered also as a system itself within its environment. It is also implied that the system in question can be itself part of a more complex system.

Systems can be closed or open. According to Bertalanffy, closed are the systems '*which are considered to be isolated from their environment'* [3]. These are systems treated by conventional physics as for example chemical reactions in a closed vessel. As Luhmann puts it, closed systems are only a 'limit case' [4]. Bertalanffy states that all living organisms are open systems [3]. He defines an open system as *'[...]a system in exchange of matter with its environment, presenting import and export, building-up and breaking-down of its material components'*.

In a closed system, the initial conditions determine a particular final state. Consequently, a change of the initial conditions results to a different final state. However, this is not the case in open systems. The notion of *equifinality* describes the property of open systems to achieve the same finals state upon different initial conditions [3]. An example in biology is the property of organisms of the same species to reach a specific final size even though they start from different sizes and going through different growth's courses.

B. Xenakis, Cybernetics and his Markovian Stochastic mechanism

It is well-known Xenakis' relation of music and mathematics. He is the first one that introduced systematically the notion of probability in music [5]. Even so, what is not so obvious is his connection of music with Systemics. In a lettre to Hermann Scherchen, in 1957, Xenakis writes: *'[...] j'ai trouvé que des transformations qui sont à la base de la cybernétique, je les ai déjà pensées et utilisées dans les Metastaseis, sans savoir alors que je faisais de la cybernétique!'* [6].[2]

In the description of his 'Markovian Stochastic Music', Xenakis explains the theory behind *Analogique A* (1958-59), for strings and *Analogique B* (1958-59), for tape [7]. He is using step-by-step the method of Ashby found in *An Introduction to Cybernetics* [8]. In particular, Xenakis shows the sonic transformations with matrixes, and as Ashby, Xenakis starts with determinate transformations continuing with stochastic transformations.

In his basic hypothesis, Xenakis claims that *'[a]ll sound is an integration of grains, of elementary sonic particles, of sonic quanta'* [7]. According to this hypothesis, it is possible to analyse and reconstruct any existing sound or even create non pre-existing sounds as a combination of thousands of grains. His so-called granular hypothesis is connected with the production of timbres, where *second order sonorities* emerge from clouds of sonic grains. As Di Scipio points out, it is possible to describe second order sonorities as a question of emerging properties of sound structures. According to Bregman:

> '*Sometimes, in the study of perception, we speak about emergent features. These are global features that arise at a higher level when information at a lower level is grouped. [...] Because nature allows structures to behave in ways that are derived from properties of parts but are unique to the larger structure, perception must calculate the properties of structures, taken as wholes, in order to provide us with valuable information about them.'* [9]

Concerning *Analogique B*, even if it may not be considered a particularly successful application of the theory, it is very significant since it is regarded as the first work of granular sound synthesis [10]. In the basis of Xenakis' application of his hypothesis, as he describes, there is a mechanism, *'the "analogue" of a stochastic process'* [7]. Xenakis explains the compositional process within his model: '*At first we argue positively by proposing and offering as evidence the existence itself; and then we confirm it negatively by opposing it with perturbatory states'* [7]. More precisely, the composer on the on hand is causing perturbations to the mechanism, while on the other hand he lets the mechanism approach the state of equilibrium. This dialectical process lets the mechanism manifest itself.

C. Di Scipio's critique on Xenakis

Di Scipio claims that the stochastic laws, which Xenakis is using to apply his hypothesis, are not capable of determining the emergence of second order sonorities [10]. He explains: '*Just as the pizzicatos of Analogique A could not but remain string pizzicatos, however dense their articulation, the electronic grains in Analogique B remain just grains and do not build up into more global auditory image.'*

Summarizing the conclusions of Di Scipio, Xenakis' mechanism: 1) is sensitive only to initial conditions, 2) its process is oriented towards a goal 3) the goal changes upon different initial conditions [10]. I have to add here that all of the above clearly show the characteristics of a 'closed system'.

In addition to the three above conclusions on the mechanism, Di Scipio also claims that in Xenakis'

[1] For a more detailed presentation of Systemics see the second chapter of [1].

[2] '[…] I think that the *transformations* which are on the basis of cybernetics, I have already though and used them in *Metastaseis* without knowing that I was doing cybernetics!'.

model: 1) timbre and form are the result of *'one and the same creative gesture'* 2) The hypothesis of second order sonorities can be successfully applied within a self-organised system's model [10]. The first point is actually an interpretation of the model that has been proven very fruitful in Di Scipio's music. However, as the application of the theory suggests, it is not an intrinsic aspect of the theory. Nevertheless, Di Scipio is using this interpretation as a fact in order to conclude that Xenakis' mechanism *'tends to establish itself a self-organising system'* [10].

D. Di Scipio's Audible Ecosystemics

Although I believe that Di Scipio's two conclusions, stated above are personal interpretations of Xenakis' model, in his personal approach they are proved particularly effective. In his model, what he calls *Audible Eco-Systemic Interface,* Di Scipio is placing the mechanism of Xenakis in an 'updated' systemic context. Here, the 'closed system' is replaced by an 'open system', a 'self-organised system'. In addition, unlike Xenakis' mechanism, all operations producing the sound result are taking place during the performance. Consequently, at the same time with an alternative approach to that of Xenakis, Di Scipio is also proposing a new interpretation of live interactive composition.

The basic concept is that the composer creates a DSP capable of self-organisation, a kind of music *organism* able to 'adapt' in a given concert's space, the organism's *environment*. The sound result depends solely on the organism's interactions with his environment, as *there is no pre-recorded material used at any point during the performance*. This adaptation is the result of the *organism's properties* causing changes to the organism's processes as a consequence of its constant communication with the given *space's properties*.

Finally, in Di Scipio's approach, the creation of sound material and of musical design are parts of one and only process [12]. As he describes, the composer lets '*global morphological properties of musical structure emerge from the local conditions in the sonic matter*' [12]. With his proposition for a *Theory of Sonological Emergence*, form becomes the formation of timbre.

E. Critique on Di Scipio

In Di Scipio's approach, the composer's focus of control is deliberately put in one and only temporal level of organisation, which is the basic micro-temporal level, letting the higher levels in favour of any occasional system's spatiotemporal dynamics. In any organisation, the control of the basic elements' formation including their interactions *does not necessarily signify the control of the formation of the whole system*. Even if there is a controlling process over the design of a system's level, that is to say the elements, their properties and their interactions, *the emerging properties of the higher organisational levels are irrelevant from this controlling process.*

Di Scipio, in favour of his persistence to *microstructural sonic design*, is giving away control of the different temporal levels' formation. Consequently, the composer is losing control over the final result while notion of formal structure is abandoned. I do not find any other reason for this persistence other than to attain continuity among the different temporal levels of organisation, since each level above is formed solely from the interactions of the level below it. Clearly, Di Scipio's approach is exclusively a *bottom-up organisation.* Nevertheless, as Mitchel explains, all adaptive systems preserve balance between bottom-up and top-down processes with an optimal balance shifting over time [13]. Di Scipio's model may be a self-organised system but its organisation lucks the multi-level processing of adaptive systems.

As a general principle of Di Scipio, the system's evolution in time is the result of its interactions in an elementary level. Nevertheless, there are certain cases in which a regulative process can be triggered and change the system's behaviour. For instance, in some works he is using a process that counts constantly the sound's activity in space. If it perceives that there is not enough activity, a set of microphones positioned in a different space are opened, feeding the ecosystem [14]. In this case, even if he designs the interactions of a higher level and he is giving again the control to the system itself, *the process causing the change of behaviour to the mechanism is not emerging from the basic elements of the system*. It is an automation that, to put it in his own words, *'is forcing the system to change from the external'* [10].[3] Notably, even if this process occurs rarely, it contradicts his theory of *microstructural design* since this process establishes differentiation in a higher organisational level and this there is an implication of *macro-structural form.* Even though, the however sporadic sequence of behavioural changes, it is not predetermined but is in question of the occasional ecosystem's dynamics.

Another conceptual contradiction is the influence of the performer over the result. In principle, the role of the performer is deliberately diminished, while it is the dialog between the system and the occasional space of the performance that creates the music. Nonetheless, in some cases, the performer makes changes to the input of the machine, which can clearly be considered as an interaction. For example, in the second work of his series *Audible Ecosystemics*, *Feedback Study* (2003), three *'gesture morphologies'* are proposed to the performer*'[a]s a general guideline'* [15] (Fig. 1).

F. Self-Organised Electroacoustic Music

With the term *self-organised music,* I refer to *the result of the interactions between some predefined structures and an occasional context of performance, through a particular interpretational model.*

Since here, our discussion is within the context of electroacoustic music, the 'predefined structures' are a particular setting of the DSP, while the 'interpretational model' is the definition of real time control parameters, what Di Scipio refers to as Control Signal Processing (CSP).

[3] Reference to Di Scipio's criticism on Xenakis' stochastic mechanism.

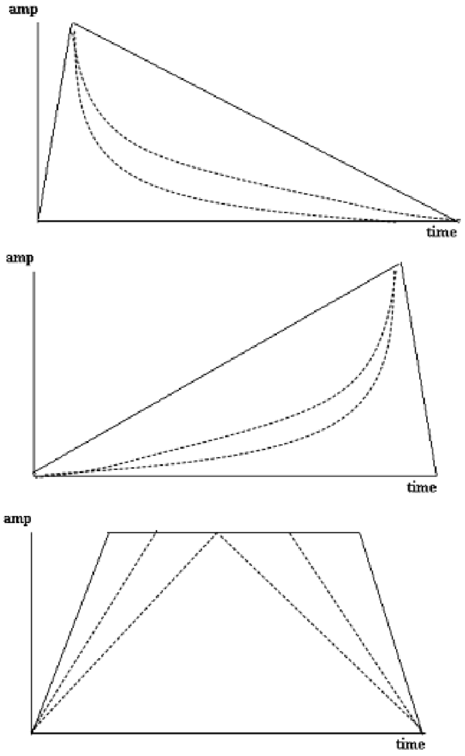


Fig. 1. The three 'gesture morphologies' over the input as guidelines to the performer of Feedback Study [15].

Based on Di Scipio's self-organised system, combined with the model of Second-Order Cybernetics, I have attempted to create a general model of self-organised electroacoustic music (Fig. 2):
The system's goal is to *control* a number of preferable variables, which represent specific sonic features. At the same time, the *perturbations* on the system are any unforeseen sounds that destabilize the system's preferable variables, in other terms *noise*. The system *observes* auditorily its environment, which is the sonic space of the performance. The process of *perception* is possible through the microphones (the sensory organs) representing the sound digitally. The representation of sound is treated in two different lines: the DSP and the CSP. Within the CSP setting, combinations of values, representing specific sonic features, influence the values of the DSP through a mapping function, which can be linear or non-linear. In this way, the DSP's characteristics are regulated from the CSP, at the same time with the DSP's processing. The result of the system's process *acts* sonically on the performance space, translated into sound through the speakers. This sonic action has an impact on the 'dynamics' of sound in space. Moreover, the *perturbations* of the environment influence sound's dynamics and indirectly destabilize the system. Finally, the circle restarts with the whole sound result in the performance space that again is perceived from the system.

III.EPHEMERON: EQUIFINALITY AND CONTROL IN SELF-ORGANISED MUSIC

I will now present my work *Ephemeron* a direct result of the research on the field of Systemics and its applications to music. Through that, I will show that it is possible to conduct emerging situations applying the systemic principle of 'equifinality'. Moreover, I will demonstrate that it is possible to acquire control over these situations and their properties in order to use formal structure over time.

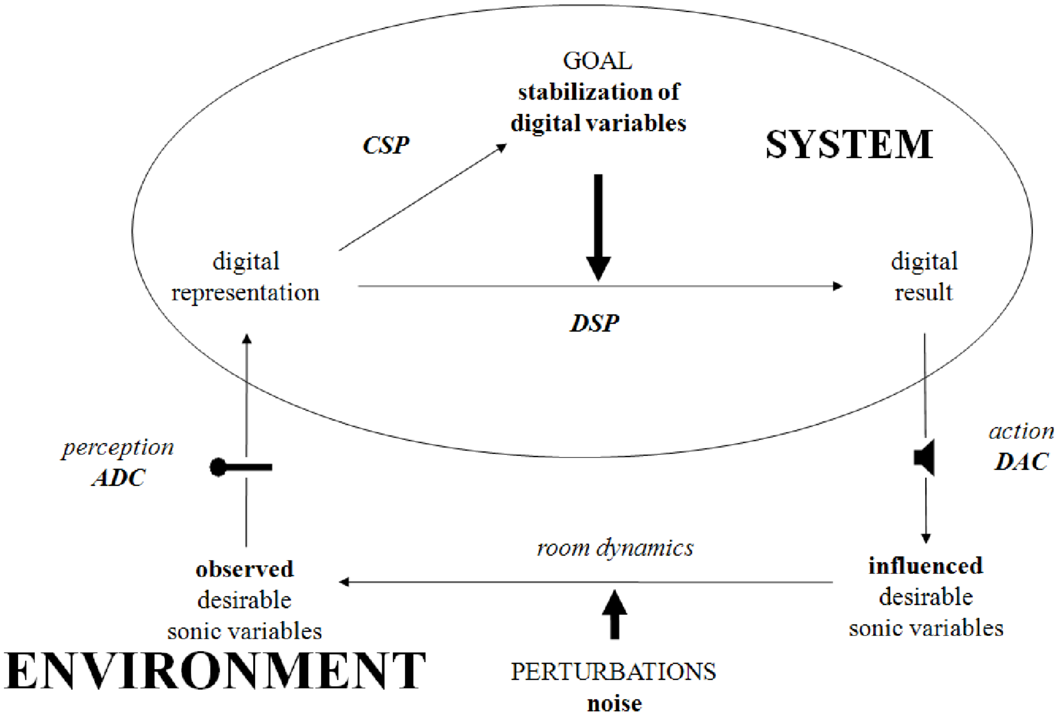


Fig. 2. A general model of self-organised electroacoustic music
(interpretation of the written and schematic description of the model of second-order cybernetics found in [16])

A. Ephemeron: The Work

Ephemeron was commissioned by Pedro Bittencourt and it was mostly developed in the Kubus concert hall of ZKM where it was also premiered. The program note provided *after* the concert was the following:

> *Microphones wide open were listening to you, listening to all of you. A newborn and constantly changing organism, existing in its unique space, was fed from every single action, every little sound of yours. Sound was flowing from the speakers manifesting the organism's existence into the concert hall. You were a unique unit of the audience with your unique perception. The audience, one entity, was fed from the organism's sounds, listening through your ears, listening through everyone's ears.*
>
> *The audience now is spread.*
>
> *The organism is no more here.*

The program note shows from the auditor's point of view the concept of the work. The work's ephemeral character is stretched and the systemic framework is implied.

A characteristic of *Ephemeron's* performance is that the sound material feeding the organism, at least at the beginning, is exclusively the applause of the audience which responds to the previous work. The organism reflects the audience's own action back to it, creating a work of music out of it.

I will describe the organism's main structure avoiding the confusing classification among global system, sub-systems and so on. For that, I will be based on the metaphor of a live organism, using biological terminology. This terminology is also coherent in the context of Systemics, giving a clear hierarchical structure.

First, an important clarification has to be made between the *'genetic' structure* of the organism and the *manifestation* of it. Staying loyal to the metaphor, we will use the distinction between the *genotype* defined as *'the sum total of the genetic information at all loci in an individual organism'* and the *phenotype*, the *'observable physical or behavioural properties of an organism that are produced by the interaction of genotype and environment during growth and development'* [17]. Here, the 'genotype' is the electronic algorithm along with the speakers and microphones. On the other hand, the organism's 'phenotype' is the sonic manifestation in a particular spatiotemporal situation. The organism's 'environment', in which the 'phenotype' results, is the actual space with its particular acoustic features, including any sound coming from the audience or from any other source.

The structure of the genotype is built in terms of control over the occasional manifestation of the organism in time but also in space. The organism's genotype has three major parts, which we will call *organs*. Each organ is consisted of four *tissues*. Finally, a tissue is formed from two *cells*, which are the basic organisational element of the phenotype.

This hierarchical structure is based on degrees of control (Fig. 3*)*. In the highest organisational level, the performer or the organism itself can influence parameters that affect all system's parts. This global parameters control the different organs, controlling the tissues, which finally control the cells.

Apart from the organism's structure in terms of control, the organism setting in space has also a specific structure according to the spatial distribution of its 'sensory organs' (the inputs, i.e. the microphones) and its outputs (the speakers). Each cell is manifested from only one speaker and it is fed from only one microphone. A

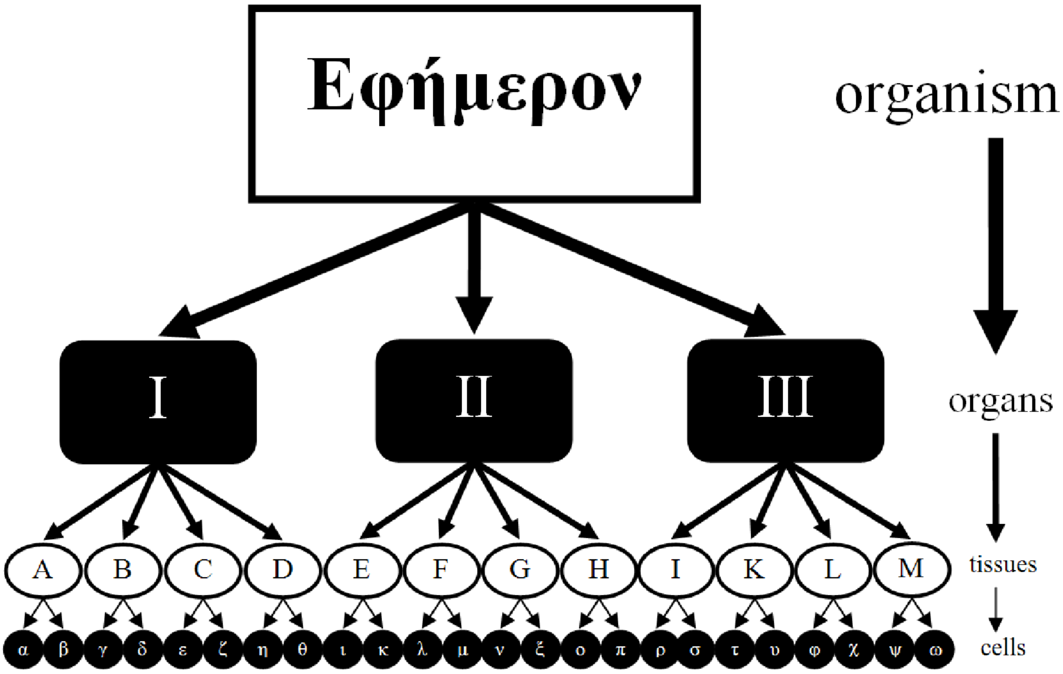


Fig. 3. The structure of the organism in terms of control's distribution.
Greek letters stands for cells, Latin letters for tissues and Latin numerals for organs.

speaker may project more than one cell. Each tissue has a unique combination of inputs and outputs.

Before the manifestation of the mechanism, between his birth and his death, its genotype has also to be adapted to the particular properties of the concert space. So far, four different 'adaptations' of *Ephemeron* have been existed in four different concert spaces. The first adaptation of *Ephemeron's* genotype was made in Z.K.M.'s Kubus, a forty-two speaker concert hall. Twelve speakers of the lower level, eight speakers of the highest level and four omnidirectional microphones were used. As it is shown in Fig. 4, the three organs were distributed in space using the front, right and left sides of the hall. The main field of the organism's spatial structure was arranged in the lower level of speakers. Furthermore, a secondary field was designed using the higher speaker's level. Except the differentiation in terms of high and low, between the two fields of manifestation, there was a difference in density of spatial projection. The higher structure was a 'folded' manifestation while the lower was an 'unfolded' one. The organism's parts could 'glissade' independently between the two fields (Fig. 5).

Regarding the existence of the music organism, I find useful the metaphor of a plant. Although the seed is not the plant itself, it contains an infinite number of possible existences of the plant. The existence of the plant may begin after a seed has entered in the appropriate environment (a fertile soil, an appropriate climate etc.) which can provide it with the appropriate amounts of energy (temperature, water, food etc.). Only then, the seed can start manifesting the existence of a plant. The plant's growth will pass through a series of states common to his species (principle of equifinality). Yet, it will show unique variation in the formation of his material structure, deriving from the interactions between its genotype and the environmental factors.

Accordingly, the music organism is something born within particular circumstances. The 'electronic' genotype includes infinite number of possible *Ephemera*. Its existence is interrelated with the beginning and the end of sound's appearance. More particularly, the organism starts to exist moments after energy is provided, by 'consuming' it. It stops existing after no more energy is left to consume and there is no more to be provided. In systemic context, the system manifests itself after the input's opening and dies after the input's closure.

In the basis of *Ephemeron*, each cell perceives and interprets the sonic reality of its environment. The cells are using dynamic control signal processing to interpret the perceptible loudness. This way there is constant change in the interpretation of the sonic reality it expresses. The basic function of the cell is to postpone the input information in a dynamic fashion while the time rate of the result's postponement is dynamically controlled from the system itself.

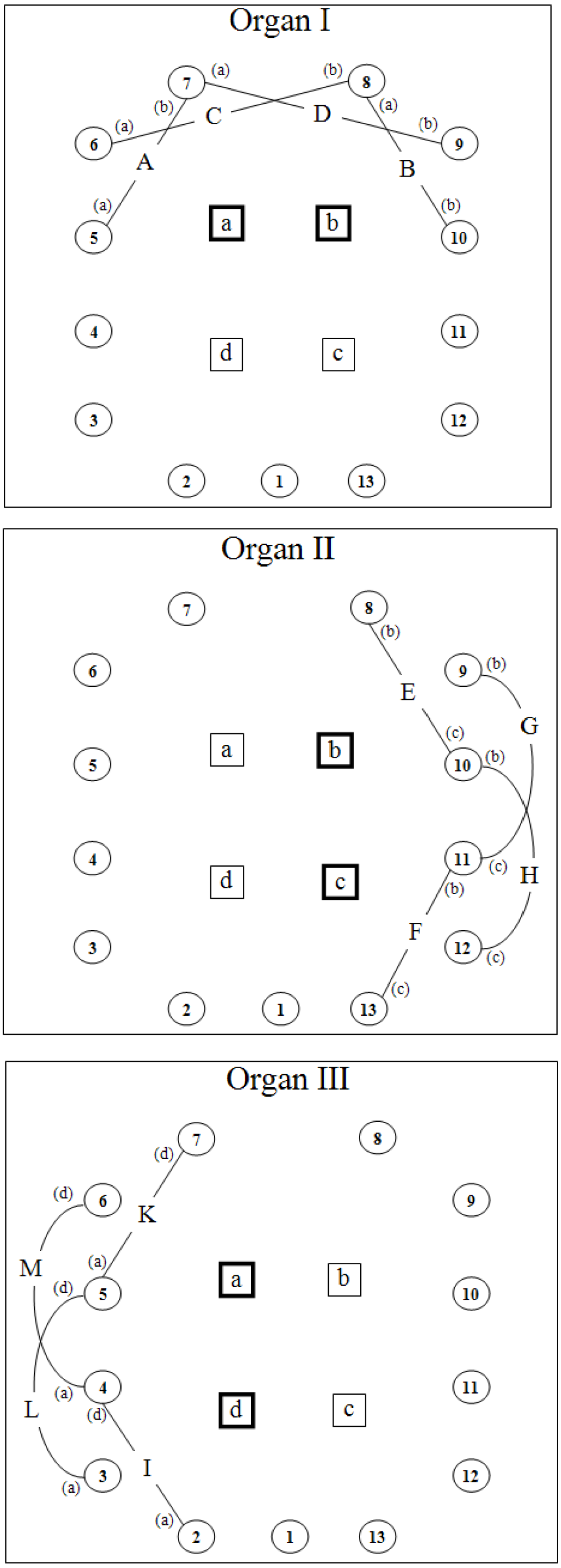


Fig. 4. The structure of the organs in terms of space. Arabic numbers stands for speakers, Latin numbers for tissues, boxed letters for microphones and letters under parentheses for the use of particular microphones within each cell

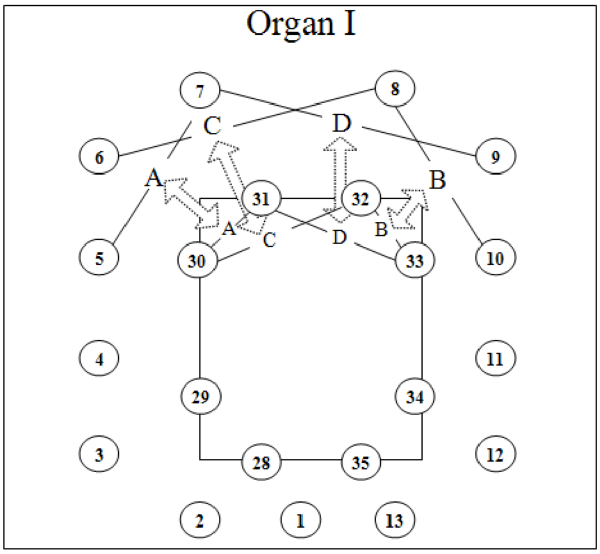


Fig. 5. The transitions of organ's manifestation between the two fields

The combination of all the cells' sonic expressions of their interpretation, make emerge something very different and much more complex that a mere reflection of the room's sonic reality. The emerging organism made by this sonic matter, it is a unique spatiotemporal expression. Spatial, since it derives from the setting of the genotype in space, and temporal, as the emergence of all ecosystemic interactions in a dynamic fashion.

B. Equifinality – Control over self-organised electronic music

My main hypothesis is that, if we consider the music organism as an open system, *it is possible to create certain conditions in which the organism will show tendency for 'equifinal' behavioural states.* As I explained before regarding 'equifinality', in an open system, '*the same final state can be reached from different initial conditions and after disturbances of the process'* [3]. I believe that *we can influence the system in order to pass from a series of behavioural states, which can be similar in any constitution of the same organism under similar circumstances.*

Consequently, in this context we are able to control the system in a basic level, by designing its elementary structures, and at the same time to acquire control over a higher organisational level, that of macrostructural form, without interrupting his ability of self-organisation. In other terms, we can let the system constitute itself, showing emerging properties over the different organisational levels and by indirectly influencing these properties we can acquire a desirable result of distinctive character. In this approach, *the composer is designing in a microstructural level and at the same time, through the role of the performer, he is controlling the sound result from a higher organisational level.*

In *Ephemeron* we have applied, I believe successfully, the above hypothesis achieving to create a 'live' organism with a specific formal constitution in time. During the concert, the organism is striving to adopt in the environment while the performer is directly changing some global parameters of the system and this way he is obstructing the system's tendency towards a state of stability. This way he is changing the organism's *behaviour*. The organism reacts by changing towards another stable state. Moreover, in each behavioural change, the information that the system perceives from its input, are interpreted in a different fashion and result to a different set of actions. The composer causes a series of changes in the behaviour of the organism.

More precisely, the performer during the concert 'interprets' a series of twenty predefined actions, causing the same number of behavioural states. His role is to 1) *change* the global parameters of the organism's structure causing a sequence of behavioural states, 2) to *monitor* the resulting changes of the organism's manifestation in time. The performer monitors the organism's manifestation in time, perceiving some expected emerging properties. He then proceeds with a new action, influencing the organism to the next change of behaviour. The performer's actions on the machine are momentary. However, each action makes the system pass through changes lasting for longer time spans. Each behavioural state is left active for a period between five and twenty seconds, according to how long it takes for the desirable properties to emerge.

In the following simple example, I demonstrate the principle in practice. The graph of Fig. 6 expresses the evolution of system's states in terms of time (thin curve). The thick curve represents the final steady states that the system states approach. As the graph suggests, the system starts with a final steady state s_5, which will be reached after time t_2. The system passes through a series of states approaching the final state s_5. Nevertheless, the performer interrupts the systems behaviour before the occurrence of s_5, by setting the new final state s_1. Again, the system changes his tendency towards the new final state s_1 that will occur in t_4. Similarly, before s_1, he changes to a new final state (s_6). This time his lets the system reach s_6 in t_5. The system *stabilizes* in s_6 and until the performer sets a new final state, there is no change to the system.

IV.CONCLUSIONS

In this paper we demonstrated some approaches of electroacoustic composition based on Systemics. First, we show the connection of Xenakis with Systemics and more particularly Cybernetics. Xenakis, with the use of his Markovian Stochastic mechanism as the basis of his model, attempts to apply the hypothesis of second order sonorities. Di Scipio argues that the stochastic laws, which Xenakis is using to apply his hypothesis, are not capable of determining the emergence of second order sonorities. Di Scipio with his model replaces the 'closed system' of Xenakis with a self-organised system. This system represents a DSP able to control its own settings in respect to the interpretation of sound's perceptual values. All processes take place during the performance, using exclusively sonic material available in the concert space.

Regarding Di Scipio's approach, I explained that the composer's focus of control is deliberately put in one and only temporal level of organisation (the basic micro-

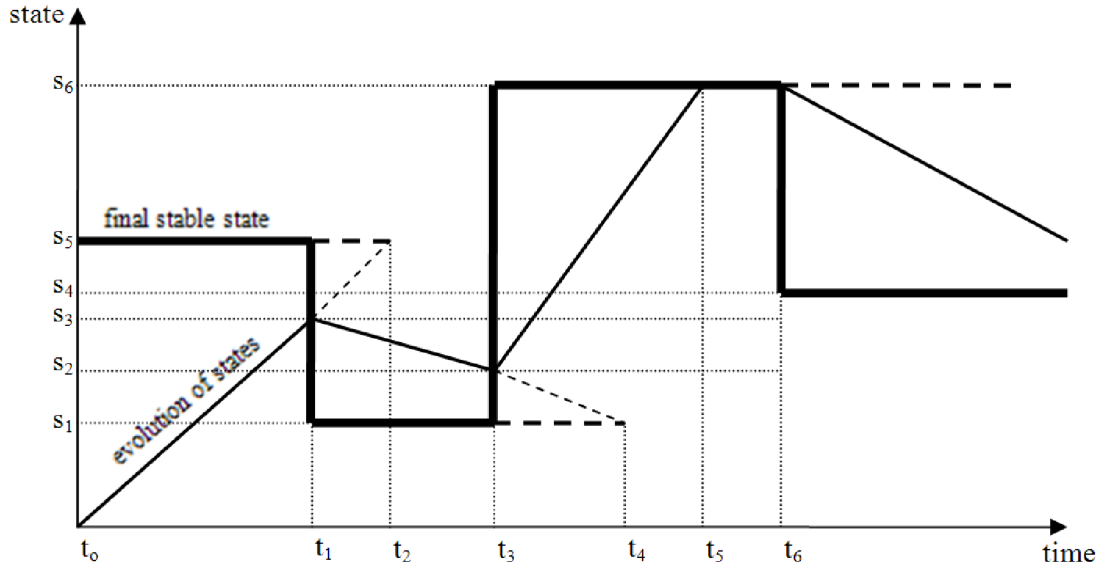


Fig. 6. The evolution of system's states

temporal level), and that this organisation is exclusively *bottom-up*. I also showed that Di Scipio's system lucks the multi-level processing which is characteristic of adaptive systems. I supposed that the reason for his persistence on designing only in a microstructural level may be to attain continuity among the different temporal levels of organisation. Also, I pointed out two conceptual contradictions in respect to his theory. The one was that, although the organisation relies only on the *microstructural sonic design*, i.e. on the basic level, there are cases where the system triggers a process that applies control over higher organisational levels. The other was that, although in principle is only the dialog between the system and the occasional space of the performance that creates the music, there are cases in which the performer makes changes to the input of the machine, establishing an interaction with it.

I defined self-organised music as the result of the interactions between some predefined structures and an occasional context of performance, through a particular interpretational model. I also attempted to create a general model of self-organised electroacoustic music, based on Di Scipio's model, interpreted through the model of Second-Order Cybernetics.

In the third section, I presented my work *Ephemeron* a self-organised system with the metaphor of a living organism. I made the distinction between its genotype and its phenotype to distinguish the 'electronic genetic code' from the manifestation of it in interaction with the environment. I formulated a hypothesis based on the systemic principle of 'equifinality and I show through the description of *Ephemeron*'s performance that it is possible to conduct emerging situations. Finally, I demonstrated that we can acquire control over these situations and their properties in order to use formal structure over time.

ACKNOWLEDGMENT

I would like to thank Horacio Vaggione for providing me with his valuable 'feedback' on the draft of this paper. I would also like to thank Agostino Di Scipio for the material and the directions he gave me throughout my Masters research and for his detailed comments on the final product. Finally, I have to acknowledge that this research owes a lot to Perdo Bittencourt for the commission and to ZKM for the performance of my work *Ephemeron*. It was this the occasion, which gave me the opportunity to develop my ideas and apply them in a creative context.

REFERENCES

[1] Ph. A. Kollias, *Systems Thinking and Music: The connections of Iannis Xenakis and Agostino Di Scipio with systems thinking. The proposal of a systemic model of symbolic music,* English version of 2008 of the original french version : *La Pensée Systémique et la Musique, les rapports de Iannis Xenakis et d'Agostino Di Scipio à la pensée systémique. La proposition d'un modèle systémique de la musique symbolique,* Unpublished MA Dissertation. Paris: Université de Paris VIII, 2007.

[2] Ph. A. Kollias, "Music and Systems Thinking: Xenakis, Di Scipio and a Systemic Model of Symbolic Music," *Proceedings of 5th Conference of Electroacoustic Music Studies Network,* Paris, 2008, in print.

[3] L. von Bertalanffy, *General System Theory: Foundation, Development, Applications*, New York: George Braziller, 1968; Revised edition, 1969; 15th paperback reprint, 2006.

[4] N. Luhmann, *Soziale Systeme: Grundriß einer allgemeinen Theorie.* Frankfurt am Main: Suhrkamp Verlag, 1984; Reference to the English edition: *Social Systems,* translated by John Bednarz, Jr. and Dirk Baecker, Stanford, California: Stanford University Press, 1995.

[5] I. Xenakis, "La crise de la musique sérielle," *Gravesaner Blätter* Vol.1 ; no.1, 1955, pp.2-4 ; Republished in: I. Xenakis, *Kéleütha*, Edited by Alain Galliari, Introduction of Benoît Gibson, Paris: L'Arche, 1994.

[6] I. Xenakis, letter to Hermann Scherchen of 19 December 1957, Archives Xenakis, National Library of France; as found in M. Solomos, 'Les «operations mentales de la composition » (Xenakis)', *Intellectica*, vol. 48-49; issue 1-2; 2008; electronic version.

[7] I. Xenakis, *Formalized music: Thought and mathematics in composition,* Harmonologia series; no. 6. New York: Pendragon Press, 1992.

[8] W.R. Ashby, *An Introduction to Cybernetics*, London: Chapman & Hall, 1956. Référence to the 2nd édition, 1957 ; online version, 1999 : http://pcp.vub.ac.be/books/IntroCyb.pdf (access: 2/3/2006)]

[9] A. S. Bregman, *Auditory Scene Analysis*, Cambridge, Massachusetts & London, England: MIT Press, 1990; Reference to the 2nd paperback edition, 1999.

[10] A. Di Scipio, "Clarification on Xenakis: The Cybernetics of Stochastic Music", *Presences of Iannis Xenakis*. Paris: CDMC, 2001, pp.71-84

[11] A. Di Scipio, "'Sound is the interface': From interactive to ecosystemic signal processing", *Organised sound: An international journal of music technology*, vol. 8; No. 3; 2003, pp.269-277

[12] A. Di Scipio, "Formal Processes of Timbre Composition, Challenging the Dualistic Paradigm of Computer Music, A study in Composition Theory (II)", *Proceedings of the ICMC International Computer Music Conference*, Aarhus, S. Francisco: ICMA, 1994; electronic version provided by the author.

[13] M. Mitchel, 'Complex Systems: Network Thinking', *Artificial Inteligence*, vol. 170, Issue 18, December 2006, pp.1194-1212; reference to the draft version; online version: http://www.santafe.edu/research/publications/workingpapers/06-10-036.pdf (accessed: 9/6/2008)

[14] A. Di Scipio. *Notes and recordings from the seminars of Agostino Di Scipio in interaction with the participating students of CCMIX between 26 and 29 March 2007;* unedited material.

[15] A. Di Scipio. *Audible Ecosystemics; n.2a, Feedback Study*, 2003; draft version provided by the composer.

[16] F. Heylighen, and C. Joslyn, 'Cybernetics and Second-Order Cybernetics', *Encyclopedia of Physical Science & Technology*, ed. R.A. Meyers, New York : Academic Press; 3rd ed., 2001 ; pp.155-170, online version: http://pespmc1.vub.ac.be/Papers/Cybernetics-EPST.pdf, (accessed: 24/7/2007)

[17] L. L. Mai, M.Y. Owl, and M. P. Kersting, *The Cambridge Dictionary of Human Biology and Evolution*, Cambridge University Press, 2005, electronic version.

An algorithm for real-time harmonic microtuning

Marc Sabat
Technische Universität/Audiokommunikation, Berlin, Germany, masa@plainsound.org

Abstract — **Subtle inflections of pitch, often performed intuitively by musicians, create a harmonically sensitive expressive intonation. As each new pitch is added to a simultaneously sounding structure, very small variations in its tuning have a substantial impact on overall harmonic comprehensibility.**

In this project, James Tenney's multidimensional lattice model of intervals ('harmonic space') and a related measure of relative consonance ('harmonic distance') are used to evaluate and optimize the clarity of sound combinations. A set of tuneable intervals, expressed as whole-number frequency ratios, forms the basis for real-time harmonic microtuning. An algorithm, which references this set, allows a computer music instrument to adjust the intonation of input frequencies based on previously sounded frequencies and several user-specified parameters (initial reference pitch, tolerance range, pitch-class scaling, prime limit).

Various applications of the algorithm are envisioned: to find relationships within components of a spectral analysis, to dynamically adjust a computer instrument to other musicians in real time, to research the tuneability of complex microtonal pitch structures. More generally, it furthers research into the processes underlying harmonic perception, and how these may lead to musical applications.

I. Introduction

In *A History of 'Consonance' and Dissonance'* James Tenney identifies several distinct conceptions of these two terms in the theory and practice of Western music. In particular, he singles out contributions made by Hermann von Helmholtz in identifying a potential *psychoacoustic* basis for defining them: namely, as properties of sound that have to do with the beating speeds of partials and combination tones in a complex structure of pitches. Helmholtz posits a relationship between consonance and the elimination of slow beats caused by unisons of common partials. Conversely, he associates dissonance with maximal roughness (partials beating between 30 and 40 Hz). Helmholtz' theory suggests that consonance, as well as *dissonance,* can be most effectively maximized by tuning sounds in Just Intonation, that is, as integer ratios of frequencies, because such sounds have a repeating (periodic) structure. Periodicity emphasizes both the sensation of stability and smooth fusion in beatless consonances as well as the regularity of intermittent pulsations causing roughness in dissonance.

Composers in the 20th century have established a basis for *musically* exploring Helmholtz' premises about harmonic perception. Arnold Schoenberg, John Cage and Edgard Varèse, among numerous others, contributed to the emancipation of dissonance and noise (inharmonic sound complexes) as acceptable musical material. Harry Partch, Ben Johnston and La Monte Young extended the late 15th Century model of Just Intonation, which was a tuning system using small-number frequency ratio intervals derived from the prime numbers 2, 3, and 5, to include relationships produced by many higher prime numbers (7, 11, 13, 17, 19, 23, 29, 31, ...).

Building on Leonhard Euler's theory of mathematically evaluating relative consonance, Tenney proposes a general multidimensional lattice model called harmonic space, in which the *harmonic distance* (HD) between two pitches comprising a musical interval may be well defined. Each interval is represented in this space by the prime factor exponents of a frequency ratio b/a, expressed in lowest terms, and measured from an arbitrarily tuned origin (1/1). This ratio (b/a) may be exactly equal to the initial interval, or it may be a nearby approximation, taking into account detuning within a specified range of *tolerance.* Harmonic distance (HD) is defined by the following equation.

$$HD = \log(ab) \tag{1}$$

One may also simply consider the lowest-terms product of numerator and denominator – the integer (ab) – which I call HD-product.

Thus, the harmonic space representation of musical intervals may be constrained by *tolerance*, by the number of prime factors included in the model (dimensionality or *prime limit),* and finally by *harmonic distance.*

II. Melodic and Harmonic Relationships

Aristoxenus, writing in the fourth century B.C., distinguishes the continuously gliding pitch changes characteristic of *speech* from the sustained, discrete tones used in *singing.* In particular, he notes that the more precisely such tones are maintained at correctly chosen pitch-heights, the more clarity melody acquires.

In the simplest sense, melody is based on three possible relations between pitches: *the same, higher* and *lower*. The property of being the 'same' allows for a certain tolerance, in the sense that small variations of pitch are not perceived as *melodically different.* Perhaps some of the finest gradations may be heard in the *srutis* of South Indian melody. In my own experience, intervals smaller than 1/6 of a tone (approximately 35¢) begin to take on the character of enharmonic shadings of pitch rather than functioning as distinct tones. Traditionally, theory has referred to such microtonal variations as *commas* and *schismas*.

On the other hand, the relations 'higher' and 'lower' tend to be divided *musically* into two general types of melodic interval: those which fall within one critical bandwidth (roughly an 8/7 ratio of frequencies) are said to move by *step,* and are referred to as *tones* and fractions thereof (e.g. semitone, quartertone, sixthtone); larger intervals are often described as *leaps.*

Claudius Ptolemy describes the situation more precisely by identifying three classes of non-unison intervals: homophones, concords, and the melodic. Homophones are manifestations of the *octave-equivalence phenomenon*. Concords are those intervals 'nearest' to the homophones, that is, the ratios 3/2 (diapente) and 4/3 (diatessaron) and their octave-composites. Melodic intervals include the difference between the primary concords (the tone 9/8) and other 'nearest' epimoric ratios of the form

$$(n+1):n \qquad (2)$$

Generalizing from this approach, we begin by examining the relations between *simultaneously* (rather than *successively)* sounded pitches. The advantage of such a method is that very slight variations of intonation, which might not disturb our perception of intervals when pitches are sounded one after the other, may be clearly distinguished when two tones are heard at the same time.

One may proceed as follows: first, by defining an extended set of concords – intervals *which may be accurately determined by perception alone;* second, by considering the set of all possible melodic movements between these intervals. This would be a repertoire of melodic steps and leaps, a complete range of *possible musical intervals,* each of which could be accurately reproduced in reference to a virtual implied third pitch.

Once again, perception suggests a rough division into three classes. In detuning one sound from unison with another, there is a region of pleasant beating, followed by a region of *roughness.* Throughout this entire range it is often difficult to hear two distinct pitches. Instead, there is a tendency to perceive a single average pitch and amplitude modulation as expressed by the well-known trigonometric identity

$$\cos u + \cos v = 2\cos((u+v)/2)\cos((u-v)/2) \qquad (3)$$

Once the distance between pitches reaches an 8/7 interval or more, the sensation of mutually generated roughness tends to be replaced by a smoother coexistence of the two separate pitches, and it becomes considerably easier to distinguish the *sonorous properties* of intervals. These variations of sound quality, which do not primarily have to do with the melodic notions 'higher' and 'lower,' may be called *harmonic* relationships. The phenomenon of octave equivalence, which allows men and women (or boys) singing at different octaves to believe that they are producing 'the same' pitch, is one clear example.

Harmonic relationships between two sounds are largely determined by beating and unison relationships between partials, as well as characteristic patterns produced between relative beating rates on different critical band regions, which are associated (in well-tuned sounds) with a sensation of *spectral fusion.* The phenomena described above in reference to the detuning of a unison are now replicated between various spectral components. Partials lock into periodic patterns, creating auditory phenomena similar to the visual patterns made by a stroboscope.

This process may be effectively investigated by constructing just intervals using additive synthesis, eliminating common partials, and hearing at what point the distinctive qualities defining specific intervals begin to disappear. To a lesser degree, harmonic relationships may also be deduced from the periodicity of a composite waveform. However, listening for beating in a non-unison interval produced between two pure sinewaves and attempting to tune it exactly is considerably more difficult than the same task with two spectrally rich tones.

Thus, in the case of simultaneously sounded pitches, discrete *steps* may most readily be found by *harmonically* examining the larger intervals, which *melodically* would be considered to be *leaps.* Those intervals that may be *precisely and repeatedly produced by perception alone* I refer to as *tuneable intervals.*

III. Tuneable Intervals

The notion of a range of tolerance within which an interval may be considered to be *de*-tuned or *mis*-tuned clearly implies that infinitely many *other* rational relationships situated in the range are not heard as distinct intervals at all. Slight detuning, producing slow *Leslie-speaker* phasing or beating in the form of *vibrato,* is often a very beautiful musical inflection of an interval that in no way damages its comprehensibility. A highly developed culture of such inflections applied within a very precisely tuned context is evident, for example, in the *gamakas* of South Indian classical music. Alternately, harmonically conceived keyboard music by J.S. Bach or Frederic Chopin, for example, is based on the ranges of tolerance inherent in the theoretically "out of tune" systems of well-temperaments or equal temperament.

At the same time, there are often *several* variations of fine-tuning possible within a given tempered interval-class, each of which may be well determined by listening for and eliminating beats. One common example is the tritone 10/7 (617.5¢) and the diminished fifth 7/5 (582.5¢), which differ by a melodic sixth-tone 50/49 (35¢). The tempered interval 600¢ found on MIDI keyboards lies in-between the two just ratios.

In addition to such variations of the well-known intervals, there are also many intervals which lie outside the range of tolerance of the 12 tempered pitch-classes but which may nevertheless be precisely heard. Most familiar in the context of Arabic music: the 'neutral' third of Zalzal (27/22 or 354.5¢).

In general, there are more intervals that can be tuned by ear than the tempered system distinguishes, but this collection of intervals is also somehow bounded. This suggests that a provisional set of *tuneable intervals* may be found. Some criteria suggest themselves for the scope of such an investigation:

(1.) **Timbre:** Each interval may be tested with both electronic additive synthesis timbres, and with acoustic instrumental sounds. When testing with electronic sounds it is useful to eliminate the clearly audible beating between common partials. In acoustic instrument sounds, the continuous variation of amplitude in spectral components reduces the effectiveness of this kind of beating as a primary tuning cue.

(2.) **Range for each interval:** It is useful to determine the range within which each interval may be tuned. It is possible that a quantitative relationship might be deduced between the individual ranges of tuneability and the register of primary difference tones and especially of the *periodicity pitch* (a virtual fundamental common to both pitches).

(3.) **Harmonic Distance:** Since intervals become increasingly difficult to hear as their ratio-numbers increase, it might be assumed that above a certain HD intervals are no longer directly tuneable, and would instead have to be constructed from simpler intervals.

(4.) **Octave Equivalence:** Some intervals may quite readily be tuned in a wide voicing (larger than an octave) whilst remaining difficult or impossible to tune in closed position. Once a set of tuneable intervals has been established, the effect of octave equivalence within this gamut may be investigated, particularly in relation to the possible formations of musical scales and modes.

(5.) **Range of Tuneability:** Due to critical band phenomena, small intervals are eventually perceived more readily as an amplitude-modulated average value than as two distinct pitches, as discussed above. Thus, there is a smallest non-unison tuneable interval. Similarly, as the intervals become larger, after a certain point the ability to perceive and differentiate harmonic phenomena becomes negligible. It is possible that the formula for harmonic distance must be appropriately modified or bounded to take these extreme conditions into better account.

(6.) **Distribution of Intervals:** In some cases, several similarly complex intervals happen to fall very close to each other, causing their harmonic qualities to be less readily distinguished (e.g. 13/9 and 16/11 are separated by 144/143, or about 12¢). Therefore, the relationship between tuneability and the distribution of ratios also merits further examination.

An ongoing empirical investigation (using acoustic and electronic sounds in various registers) has to date identified 122 23-limit tuneable intervals in the range 1/1 (0¢) to 28/1 (5769¢). The complete list of all distinct lowest-terms ratios between the first 28 partials, together with information about prime limit, melodic distance, harmonic distance, and an empiric evaluation of tuneability on a four-level scale (impossible, very difficult, average, easy; numerically represented as 0–3) may be found as an Appendix below.

These harmonic relationships have been used to construct the provisional form of the tuning algorithm described in this paper.

IV. Tuneable Melodic Steps

When the 122 tuneable intervals are used to define a set of pitches both above and below a given starting note, and the doubled occurrence of the unison 1/1 is eliminated, 243 distinct pitches remain. By generating a 2-dimensional 243x243 array, all of the intervals spanning these pitches may be determined. Once duplicates have been filtered, this produces a list of 3997 unique *tuneable melodic intervals* ranging from a unison (1/1 or 0¢) to just under 10 octaves (784/1 or 11537.7¢). Each of these melodic intervals implies a *virtual third pitch,* which (if sustained while successively sounding the tones forming the melodic interval) will form tuneable intervals to both initial pitches.

The fine-tuning algorithm described below uses this array of intervals as a lookup table, based on the premise that it encompasses tuneable intervals "one-step-removed" and may therefore effectively model how our brains could deduce possible connections between pitches.

A few statistical observations about this list are tabulated to indicate how finely it resolves the glissando-continuum:

TABLE I.
Tuneable Melodic Intervals by Octave

Octave	# of Intervals	Mean Step	Smallest Step	Largest Step
all	3997	2.89¢	0.07¢	62.96¢
1	785	1.52¢	0.07¢	8.34¢
2	720	1.67¢	0.07¢	8.34¢
3	667	1.80¢	0.07¢	8.34¢
4	600	2.00¢	0.07¢	8.34¢
5	431	2.78¢	0.14¢	11.35¢
6	310	3.87¢	0.36¢	13.65¢
7	225	5.29¢	0.40¢	23.34¢
8	133	9.09¢	0.79¢	28.27¢
9+	126	15.38¢	2.38¢	62.96¢

In future development, I plan to revise this array as follows: rather than building it from an empirically determined tuneable interval set, I would prefer to find a quantitative HD value above which no more tuneable intervals may be found. Then, all possible ratios found to lie within this limiting value might further be constrained by an overall range of tuneability (see discussion above), and by ranges found for each individual interval.

Based on the particular experience and intervallic discrimination of a given listener, it would then be possible to dynamically adjust the array by specifying a limiting HD and several range-related parameters.

V. Triads

So far, the discussion has been limited to the case of intonation in dyads. Clearly the situation increases in complexity with each additional note comprising an aggregate of pitches. At the same time, the most important qualitative (musical) differences are represented in the three simplest cases: *melodic intervals* (two successive pitches); *harmonic intervals* (two simultaneous pitches); *chords* (three or more simultaneous pitches).

At this point, I limit my discussion of chords to the case of triads. In the future development of the algorithm proposed here, I anticipate that it will be possible to generalize to structures involving more pitches. Eventually, such study might also be connected to the more statistical properties of large microtonal *aggregates* explored by Iannis Xenakis and by some of the spectralist composers (Gérard Grisey, Tristan Murail, Oliver Schneller). It will also require work to understand how larger aggregates form modal sets, which become clearly defined in memory even when not all pitches are sounding.

In analyzing the problem of fine-tuning an arbitrarily formed microtonal triad, it is useful to consider that there are not only three different pitches, but also *three intervals* formed between *pairs* of pitches. Taking the three pitches in ascending order (expressed as frequencies)

$$F1 < F2 < F3 \tag{4}$$

the three intervals may be summarized mathematically as

$$(F3/F1) = (F3/F2)(F2/F1) \tag{5}$$

Looking at the triad expressed in this way, and comparing the three dyads to the previously discussed set of tuneable intervals, four possibilities emerge. *In any triad, either none, one, two, or all three of the component dyads form tuneable intervals.* In the 'two' case one might note two sub-classes, depending on whether the *outer* interval is tuneable or not.

Since the composite sound of a triad is, in some way, a superimposition of three interference patterns between pitch-pairs, it seems intuitively correct to postulate that triads made up from two or three tuneable dyads will generally sound more stable and comprehensible than those with none or only one. Thus, assuming that the purpose of fine-tuning is to maximize clarity and variety within a large range of possible sound structures, any algorithm that *favors the possibility of two or more tuneable interval relationships* will produce results with a pronounced acoustical advantage.

Each triad made up from the same dyads has two possible symmetric forms, which (following Partch) I distinguish as *otonal* or *utonal.*

In the case of dyads, this may be imagined as the difference between taking an interval upward or downward from a given starting pitch. The resulting sound structure is related by a transposition of register, but otherwise exactly identical.

In the case of three pitches, written as above, the two smaller intervals F2/F1 and F3/F2, taken together, "add up" to form the outer interval F3/F1. It is possible to reverse the order of these two smaller intervals by defining a new pitch

$$F2^* = (F3/F2) \cdot F1 \tag{6}$$

(So F2*/F1 = F3/F2 and F3/F2* = F2/F1.)

If the two chord structures F1:F2:F3 and F1:F2*:F3 are expressed in lowest terms, either they will both turn out to be identical (if F2* = F2, i.e. F3/F2 = F2/F1) or one of these two forms may be found to consist of "smaller" numbers. In this case, define "smaller" in the following sense: if the outer terms of both forms are the same, take the triad with a smaller middle term; if not, take the one with a smaller first term. This I would then call the *otonal* form.

To briefly explain this idea, take as an example the relationship between a major and minor triad. Both are composed of a major third (5/4) and a minor third (6/5) adding up to a perfect fifth (3/2). As chords, they may be represented by the frequency ratios 4:5:6 (major) and 10:12:15 (minor). The numbers indicate that the minor triad occurs later in a harmonic (overtone) series than the major triad, which is therefore acoustically simpler (easier for the ear to analyze) and which, according to the definition above, is *otonal.*

Consider the property that any tuned aggregate of pitches shares not only a *common fundamental frequency* but also a *least common partial.* It is thus also possible to express chord proportions in relation to their common partial. Still thinking of overtone series, in the case of both 4:5:6 and 10:12:15 the least common partial will be 60 (the least common multiple in both cases). Thus, *in relation to this common partial* (which, in a well-tuned triad, may be acoustically perceived as part of the composite sound) the minor triad may be expressed as 1/6:1/5:1/4 and the major as 1/15:1/12:1/10.

So, considered downward *(utonally),* the minor triad takes smaller numbers. Namely, when building a subharmonic (undertone) series downward from the common partial, the minor triad will occur sooner than the major, and so I refer to it as *utonal.*

The symmetry inherent in this argument has been compelling to many music theorists, including Rameau, Riemann and Partch, particularly as a way of "explaining" the minor triad (in spite of its dissonant difference tones) and also as a way of generalizing major-minor tonality. However, the perception of chord "stability" (which tonality requires) is based on a psychoacoustic sensation of fusion produced by harmonic spectra. Utonal sounds are thus by definition less stable than otonal sounds (because they are further away from their fundamental periodicity pitch). As utonal structures become increasingly complex, their conceptual symmetry to otonal counterparts is no longer acoustically perceptible.

Nevertheless, it is certainly possible and musically fruitful to investigate a list of simpler triads in which this quality *is* maintained. One interesting such example is the septimal triad 6:7:8 and its utonal counterpart 21:24:28.

VI. Development of the Algorithm

The problem might be summarized as follows: given two frequencies extracted from a spectral aggregate (e.g. a woodwind multiphonic timbre), what might be an effective method for extrapolating to harmonically interesting third pitches?

First, there is the problem of deducing potentially *perceptible* harmonic relationships (ratios) between the two extracted pitches. Then, once one of these ratios has been selected, the choice of a third pitch is evaluated based on the intervals it forms with both initial pitches, as well as by the overall sonorous qualities of the triad all three generate.

Imagine the three pitches forming a triangle. By slightly adjusting each vertex it is possible to find a proportionally ideal structure. The first vertex need only be adjusted if there is an external tuning standard in place (for example, A-440). The second vertex is adjusted to the first, producing a tuneable melodic step, which may or may not itself be a tuneable interval. Then any number of tuned possibilities exist for the third vertex, at least some of which produce two tuneable intervals to both initial pitches.

This model led to the current harmonic microtuning algorithm, which exists in the form of an external object, programmed in C, and compiled to run in the MaxMSP environment. This external has been implemented within a Max patch, which allows for the fine-tuning of up to three pitches, in relation to each other and to a reference frequency, and according to three additional user-specified parameters: tolerance, pitch-class scaling, and prime limit.

The program uses the list of 3997 tuneable melodic intervals as a lookup-table, searching to find the nearest possible match within a desired prime limit. If this nearest result falls outside of a user-specified tolerance range, then it is immediately output. Otherwise the program searches for the *simplest* result within the desired range, evaluated by minimizing a harmonic-distance sum.

This sum is weighted by the choice of a pitch-class scaling value, which serves to either favor the simplicity of the sounding interval (value 0), the spelling of the microtonal pitch within a notated system (value -1), or the pitch-class relationship to its reference pitch (value 1). Intermediate values produce a linear interpolation between the evaluated harmonic distances. Pitch-class HDs are computed by ignoring the octave prime dimension 2 (factoring out all the 2's in the ratios). The resulting harmonic space is called *projection space.*

Input to the external is distributed between ten available 'inlets.' Following the right-to-left logic of Max, inlet 10 accepts integers and determines a prime limit ranging from Pythagorean intonation (intervals generated by the numbers 2 and 3 and their powers) to the limits of tuneability (intervals made from the primes 3, 5, 7, 11, 13, 17, 19, 23). In future implementations, it would be interesting to have an on-off choice for each prime, allowing, for example, a set of intervals generated by the primes 3 and 7 only.

Inlet 9 accepts pitch-class scaling, in the form of a float value between -1 and 1. If 0 is entered, the algorithm evaluates intervals as they sound, without considering pitch-class. Thus, if a G, a minor seventh above A, is input as the first frequency to be tuned, and the tolerance range and prime limit allow it, the algorithm will prefer 9/5 (raising the G by a comma) to 16/9 (Pythagorean G). However, if the same pitch class is entered two octaves lower, as a G that is a major ninth below A, then the algorithm would prefer 4/9 (Pythagorean G). In both cases, the simplest *sonority* is chosen.

Sometimes, however, it might be preferable for the algorithm to choose in the manner of traditional modes and scales, repeating identically in each octave. (In this case it should retune *consistently* whenever it receives a G.) The decision may be weighted in favor of pitch class relationships to the reference frequency (values between 0 and 1), or in favor of the microtonal spellings using the Extended Helmholtz-Ellis JI Pitch Notation (values between 0 and -1). This second choice might produce slightly more complex intervals but has the advantage of keeping the notated pitches simpler, useful when dealing with a written score or whilst algorithmically generating pitches to be written down and played by other instruments.

Inlet 7 is the reference frequency (F0), expressed as a float value in Hz. I generally leave this at the standard tuning reference (440 Hz), but it is possible to use any frequency (for example, any tempered pitch in any octave).

The first six (of ten total) inlets are taken up (pairwise) with the three frequencies (F3, F2, F1), which may each be input in one of three possible numerical representations. In the following description, the terms 'inlet 1' and 'inlet 2' are used an as example: inlets 3&4 and 5&6 may be imagined as behaving identically.

If an **exact just intonation ratio** to F0 is desired, namely a pitch, *which is already tuned and ought to be left alone,* then inlet 1 and inlet 2 must both receive nonzero positive integers. If an **absolute frequency value** is to be interpreted, inlet 2 should receive a 0 and inlet 1 a positive float value.

If the input desired is **MIDI+cents** (which offers a greater degree of precision) then inlet 1 must receive a negative value, which may be calculated from the desired MIDI note. (The MIDI value of the note is multiplied by -1 and added to -1000, an arbitrary offset value, which must also be adjusted by the distance of the tempered reference frequency from A4 MIDI value 69.) In this case inlet 2 accepts positive or negative float values for deviation from the MIDI value in cents.

Each incoming frequency is tuned in relation to a *reference:* F1 is tuned to F0, F2 is tuned to F1, and F3 is tuned to either F2 or F1, whichever possibility offers the best overall result. The ability to input a ratio allows the user to specify complex just intonation intervals, which will not be retuned, as components of the structure.

VII. Current Implementation (Micromælodeon I)

The current algorithm is limited to making decisions about the fine-tuning of one, two or three pitches, received successively, sounding simultaneously. It has been implemented as the core of a virtual instrument called Micromælodeon I. Up to three sounds may be generated and selectively fine-tuned, using a simple wave-shape-morphing synthesis method combined with a filtergraph used to simulate resonances and formants.

It is possible that the three-pitch 'triangulation' process described in this paper may be effectively generalized to tune structures of up to twelve pitches related in a network as follows:

(1.) F0, F1, F2, F3 function as described above.

(2.) F4 is tuned to F0, F1, and F3.

(3.) F5 is tuned to F0, F2, and F3.

(4.) F4 and F5, combined respectively with each of F1, F2, F3 (with reference F0) produce the next six pitches – F6 through F11.

(5.) F4 and F5 (with reference F0) produce F12.

A second instrument to investigate this extended model is being planned and programmed, to be followed by empirical testing of the algorithm's performance with trained musicians, the design of hardware control interfaces for virtual instruments and implementation of a learning memory which would facilitate developing harmonic decision-making over larger musical time-scales (phrase, section, entire piece).

I anticipate that developing this 'memory' will continue my previous work with 'crystal growth' algorithms in harmonic space, which enable stochastic generation of harmonically compact pitch-clusters.

VIII. Conclusion

In informal testing, given appropriately chosen reference pitches and parameters, the Micromælodeon is readily able to find classical sets of pitches: among others, the major and minor scales tuned in Just Intonation (Ptolemaic tense [syntonon] diatonic); the 7-limit srutis used in Indian music. At the same time, it suggests subtle fine-tunings of more 'dissonant' equal-tempered chords, based on relationships of higher partials, revealing complex tonal relationships underlying 'atonal' sounds.

It is hoped that the algorithm presented here will provide a foundation for new electronic instruments, which allow for precise *musical* investigations of the phenomena of harmonic perception, by implementing well-formed *descriptive* (rather than pro- or pre-scriptive) principles of those relations between pitches which do not have to do exclusively with "higher" and "lower" (i.e. *harmony*).

APPENDIX: TABLE OF TUNEABLE INTERVALS

The intervals below are taken from the first 28 partials, ordered by rising HD-product (numerator multiplied by denominator). The complete list of unique lowest terms ratios was truncated at the point after which no more tuneable intervals were found. Boldface font indicates extensions of the original three-octave tuneable set. Italic font indicates intervals that were not found to be tuneable. The informal 'degree' column shows, on a scale from 0 to 3, my own assessment of how difficult the tuning task was using acoustic stringed instruments (violin, viola, violoncello, contrabass).

TABLE II.
TUNEABILITY OF INTERVALS SORTED BY HD-PRODUCT

Ratio (num)	Ratio (den)	Degree (0–3)	HD-product	Prime Limit	Size (cents)
1	1	3	1	1	0
2	1	3	2	2	1200
3	1	3	3	3	1901.955001
4	1	3	4	2	2400
5	1	3	5	5	2786.313714
3	2	3	6	3	701.9550009
6	1	3	6	3	3101.955001
7	1	3	7	7	3368.825906
8	1	3	8	2	3600
9	**1**	**3**	9	**3**	**3803.910002**
5	2	3	10	5	1586.313714
10	**1**	**3**	10	**5**	**3986.313714**
11	**1**	**3**	11	**11**	**4151.317942**
4	3	3	12	3	498.0449991
12	**1**	**3**	12	**3**	**4301.955001**
13	**1**	**3**	13	**13**	**4440.527662**
7	2	3	14	7	2168.825906
14	**1**	**2**	14	**7**	**4568.825906**
5	3	3	15	5	884.358713
15	**1**	**2**	15	**5**	**4688.268715**
16	**1**	**2**	16	**2**	**4800**
17	**1**	**1**	17	**17**	**4904.95541**
9	2	3	18	3	2603.910002
18	**1**	**1**	18	**3**	**5003.910002**
19	**1**	**1**	19	**19**	**5097.513016**
5	4	3	20	5	386.3137139
20	**1**	**1**	20	**5**	**5186.313714**
7	3	3	21	7	1466.870906
21	**1**	**1**	21	**7**	**5270.780907**
11	2	3	22	11	2951.317942
22	**1**	**1**	22	**11**	**5351.317942**
23	**1**	**1**	23	**23**	**5428.274347**
8	3	3	24	3	1698.044999
24	**1**	**1**	24	**3**	**5501.955001**
25	**1**	**1**	25	**5**	**5572.627428**
13	2	3	26	13	3240.527662
26	**1**	**1**	26	**13**	**5640.527662**
27	**1**	**1**	27	**3**	**5705.865003**
7	4	3	28	7	968.8259065
28	**1**	**1**	28	**7**	**5768.825906**
6	5	3	30	5	315.641287
10	3	3	30	5	2084.358713
15	2	3	30	5	3488.268715
11	3	3	33	11	2249.362941
17	**2**	**2**	34	**17**	**3704.95541**
7	5	3	35	7	582.5121926
9	4	3	36	3	1403.910002
19	**2**	**2**	38	**19**	**3897.513016**
13	3	3	39	13	2538.572661
8	5	3	40	5	813.6862861
7	6	3	42	7	266.8709056
14	3	3	42	7	2666.870906
21	**2**	**2**	42	**7**	**4070.780907**
11	4	3	44	11	1751.317942
9	5	3	45	5	1017.596288
23	**2**	**2**	46	**23**	**4228.274347**
16	3	3	48	3	2898.044999
25	**2**	**2**	50	**5**	**4372.627428**
17	3	3	51	17	3003.000409
13	4	3	52	13	2040.527662
27	**2**	**1**	54	**3**	**4505.865003**
11	5	2	55	11	1365.004228
8	7	2	56	7	231.1740935
19	3	3	57	19	3195.558015
12	5	3	60	5	1515.641287
15	4	3	60	5	2288.268715
20	3	3	60	5	3284.358713
9	7	3	63	7	435.0840953
13	5	2	65	13	1654.213948
11	6	2	66	11	1049.362941
22	3	3	66	11	3449.362941
17	4	2	68	17	2504.95541
23	3	3	69	23	3526.319346
10	7	2	70	7	617.4878074
14	5	3	70	7	1782.512193
9	**8**	**1**	**72**	**3**	**203.9100017**
25	**3**	**1**	**75**	**5**	**3670.672427**
19	4	2	76	19	2697.513016
11	7	2	77	11	782.4920359
13	6	2	78	13	1338.572661
26	**3**	**2**	78	**13**	**3738.572661**
16	5	3	80	5	2013.686286
12	7	2	84	7	933.1290944
21	4	2	84	7	2870.780907
28	**3**	**1**	84	**7**	**3866.870906**
17	5	2	85	17	2118.641696
11	8	1	88	11	551.3179424
10	*9*	*0*	*90*	*5*	*182.4037121*
18	5	3	90	5	2217.596288
13	7	2	91	13	1071.701755
23	4	2	92	23	3028.274347
19	5	2	95	19	2311.199302
11	9	1	99	11	347.4079406
25	4	2	100	5	3172.627428
17	6	2	102	17	1803.000409
13	8	2	104	13	840.5276618
15	*7*	*0*	*105*	*7*	*1319.442808*
21	5	1	105	7	2484.467193
27	4	1	108	3	3305.865003
11	*10*	*0*	*110*	*11*	*165.0042285*
22	5	1	110	11	2565.004228

16	*7*	*0*	*112*	*7*	*1431.174094*
19	6	1	114	19	1995.558015
23	5	2	115	23	2641.960633
13	9	1	117	13	636.61766
17	7	1	119	17	1536.129503
15	8	1	120	5	1088.268715
24	5	2	120	5	2715.641287
14	9	1	126	7	764.9159047
18	7	2	126	7	1635.084095
13	10	2	130	13	454.2139479
26	5	1	130	13	2854.213948
12	*11*	*0*	*132*	*11*	*150.6370585*
19	*7*	*0*	*133*	*19*	*1728.68711*
27	*5*	*0*	*135*	*5*	*2919.551289*
17	*8*	*0*	*136*	*17*	*1304.95541*
23	6	1	138	23	2326.319346
20	7	2	140	7	1817.487807
28	5	2	140	7	2982.512193
13	*11*	*0*	*143*	*13*	*289.2097194*
16	*9*	*0*	*144*	*3*	*996.0899983*
25	6	1	150	5	2470.672427
19	8	1	152	19	1497.513016
17	*9*	*0*	*153*	*17*	*1101.045408*
14	*11*	*0*	*154*	*11*	*417.5079641*
22	7	1	154	11	1982.492036
13	*12*	*0*	*156*	*13*	*138.5726609*
23	*7*	*0*	*161*	*23*	*2059.448441*
15	*11*	*0*	*165*	*11*	*536.9507724*
21	*8*	*0*	*168*	*7*	*1670.780907*
24	7	1	168	7	2133.129094
17	*10*	*0*	*170*	*17*	*918.6416956*
19	*9*	*0*	*171*	*19*	*1293.603014*
25	*7*	*0*	*175*	*7*	*2203.801521*
16	11	1	176	11	648.6820576
20	*9*	*0*	*180*	*5*	*1382.403712*
14	*13*	*0*	*182*	*13*	*128.2982447*
26	*7*	*0*	*182*	*13*	*2271.701755*
23	8	2	184	23	1828.274347
17	*11*	*0*	*187*	*17*	*753.6374671*
27	7	1	189	7	2337.039096
19	*10*	*0*	*190*	*19*	*1111.199302*
15	*13*	*0*	*195*	*13*	*247.741053*
18	*11*	*0*	*198*	*11*	*852.5920594*
22	*9*	*0*	*198*	*11*	*1547.407941*
25	8	1	200	5	1972.627428
17	*12*	*0*	*204*	*17*	*603.0004086*
23	*9*	*0*	*207*	*23*	*1624.364346*
16	*13*	*0*	*208*	*13*	*359.4723382*
19	*11*	*0*	*209*	*19*	*946.1950738*
15	*14*	*0*	*210*	*7*	*119.4428083*
21	*10*	*0*	*210*	*7*	*1284.467193*
27	8	1	216	3	2105.865003
20	*11*	*0*	*220*	*11*	*1034.995772*
17	*13*	*0*	*221*	*17*	*464.4277477*
25	*9*	*0*	*225*	*5*	*1768.717426*
19	*12*	*0*	*228*	*19*	*795.5580153*
23	*10*	*0*	*230*	*23*	*1441.960633*
21	*11*	*0*	*231*	*11*	*1119.462965*
18	*13*	*0*	*234*	*13*	*563.38234*
26	*9*	*0*	*234*	*13*	*1836.61766*
17	*14*	*0*	*238*	*17*	*336.129503*
16	*15*	*0*	*240*	*5*	*111.7312853*
19	*13*	*0*	*247*	*19*	*656.9853544*
28	9	1	252	7	1964.915905
23	*11*	*0*	*253*	*23*	*1276.956405*
17	*15*	*0*	*255*	*17*	*216.6866948*
20	*13*	*0*	*260*	*13*	*745.7860521*
24	*11*	*0*	*264*	*11*	*1350.637059*
19	*14*	*0*	*266*	*19*	*528.6871097*
27	*10*	*0*	*270*	*5*	*1719.551289*
17	*16*	*0*	*272*	*17*	*104.9554095*
21	*13*	*0*	*273*	*13*	*830.2532456*
25	*11*	*0*	*275*	*11*	*1421.309485*
23	12	1	276	23	1126.319346

ACKNOWLEDGMENT

Special thanks to the colleagues who have helped in the ongoing development of this research, especially James Tenney, Wolfgang von Schweinitz, Walter Zimmermann, Stefan Bartling, Natalie Pfeiffer and Chiyoko Szlavnics.

A particular acknowledgment is due to Rama Gottfried, whose pointed questions in my seminars at the Universität der Künste Berlin stimulated the programming of this algorithm and to Dr. Stefan Weinzierl, for his continued support of my work.

REFERENCES

[1] Albert S. Bregman, *Auditory Scene Analysis.* Cambridge: The MIT Press, 1990.

[2] Ivor Darreg, *Elastic Tuning.* Self-published, sonic-arts.org, 1988.

[3] Leonhard Euler, *Tentamen Novae Theoriae Musicae,* translated by Charles Samuel Smith (doctoral dissertation). Bloomington: Indiana University, 1960.

[4] Yonatan I. Fishman, Igor O. Volkov, M. Daniel Noh, P. Charles Garell, Hans Bakken, Joseph C. Arezzo, Matthew A. Howard, and Mitchell Steinschneider, "Consonance and Dissonance of Musical Chords: Neural Correlates in Auditory Cortex of Monkeys and Humans," in *Journal of Neurophysiology, Vol. 86, December 2001,* pp. 2761–2788.

[5] Hermann L.F. Helmholtz, *On the Sensations of Tone as a Physiological Basis for the Theory of Music,* translated (1885) by Alexander J. Ellis. New York: Dover, 1954.

[6] Ben Johnston, *Maximum Clarity and Other Writings on Music* (edited by Bob Gilmore). Urbana/Chicago: University of Illinois Press, 2006.

[7] Walter Mohrlok, *The Hermode Tuning System.* Self-published, 2003.

[8] Harry Partch, *Genesis of a Music.* Madison: The University of Wisconsin Press, 1949.

[9] Richard Parncutt, *Harmony: A Psychoacoustic Approach.* Berlin: Springer-Verlag, 1989.

[10] Marc Sabat, "The Extended Helmholtz-Ellis JI Pitch Notation," in *Mikrotöne und Mehr.* Hamburg: von Bockel Verlag, 2005, pp. 315–331.

[11] Marc Sabat, "Three Crystal Growth Algorithms in 23-limit Constrained Harmonic Space," in *Contemporary Music Review, Vol. 27. No. 1, Feb. 2008,* pp. 57–78.

[12] Arnold Schoenberg, *Theory of Harmony,* translated by Roy E. Carter. London: Faber and Faber, 1978.

[13] William A. Sethares, "Adaptive tunings for music scales," in *Journal of the Acoustical Society of America, 96 (1) July 1994,* pp. 10–18.

[14] James Tenney, *A History of 'Consonance' and Dissonance.'* New York: Excelsior, 1988.

[15] James Tenney, "John Cage and the Theory of Harmony," in *Soundings 13: The Music of James Tenney*. Santa Fe: Soundings Press, 1984, pp.55–83.

[16] Richard M. Warren, *Auditory Perception: A New Analysis and Synthesis.* Cambridge: Cambridge University Press, 1999.

NOVARS RESEARCH CENTRE, UNIVERSITY OF MANCHESTER, UK. STUDIO REPORT

Ricardo Climent [*], David Berezan [†] and Andrew Davison [*]

[*†] Novars Research Centre, School of Arts Histories and Cultures, University of Manchester, UK,

[*] ricardo.climent@manchester.ac.uk, [†] david.berezan@manchester.ac.uk, [*] andrew.davison@manchester.ac.uk

Abstract — **NOVARS is a new Research Centre started in March 2007 with specialisms in areas of Electroacoustic Composition, Performance and Sound-Art. The Centre is capitalising on the success of Music at the University of Manchester with the expansion of its existing research programme in Electroacoustic Composition with a new £2.2 million investment in a cutting-edgev new studios infrastructure. This studio report covers key aspects of architectural and acoustic design of the Studios, functionality and existing technology.**

I. INTRODUCTION

The Studios were constructed by Harry Fairclough Construction Ltd, after a cutting-edge design by Cruickshank and Seward architects and acoustic design by ARUP consulting engineers. The sound system design and installation was provided by DACS-AUDIO and the project was managed by Stuart Lockwood, Estates Capital Projects Group at the University of Manchester and coordinated by David Berezan, Director of the Electroacoustic Studios at the Martin Harris Centre for Music and Drama and MANTIS.

The new Studios facilities were officially opened in November 2007 and incorporate three large postgraduate research studios (featuring 32-loudspeaker monitoring for sound diffusion performance research and multi-channel composition, 5.1-surround and 10-channel composition environments), a fourth multi-function studio for undergraduate teaching and work, a large 14-workstation computer cluster for teaching and student compositional work, supporting technical spaces, offices and computer/ network hubrooms.

II. BACKGROUND

NOVARS is named to reference and celebrate the seminal work by Francis Dhomont (Novars). In his own words 'a reversed version of Ars Nova' - New Art, New Science. Staff and postgraduate student research areas range from acousmatic composition to machine musicianship, sound spatialisation, performance practice, live interactive systems and cross-disciplinary projects.

In education areas, NOVARS is supporting and reinforcing an existing well-integrated music pathway at the University of Manchester, both at Postgraduate and Undergraduate levels, merging areas of electroacoustic composition, instrumental composition and music theatre; including experimental and contemporary performance practice. An added value to the NOVARS research focus is the strength of the School's performance programme and the joint pathway between Music and the RCNM (Royal Northern College of Music). Accessibility to high-class performers in residence makes the environment extremely appealing for composers willing to experiment on extended techniques, chamber groups or large scale instrumental forces in combination with new music technologies.

III. STUDIO DETAILS

A. General Design Features

The massing of the elevational treatment of the building is in direct response to Manchester City Planners desire

to improve the visual expression along Bridgeford Street, where the studios were built. The variety of contrasting materials that have been used separate the different functional and circulation elements of the building, and improve the visual form of the frontage.

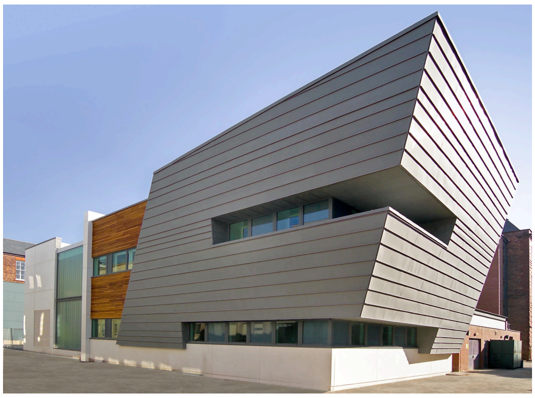


Figure 1. Façade of the building hosting the NOVARS Research Centre

The corner of the building is also a key site feature, further enhancing the external appearance of the urban block. A sharp angular corner was incorporated with a pronounced overhang and the corner is also accentuated by steeply raking the zinc cladding out in this direction.

B. Acoustic Considerations by ARUP

To reduce the effect of noise contamination both to the isolated internal studios and to the adjoining street and neighbouring buildings, the air handling plant is attenuated and contained within the ground floor plant room with no externally mounted condensers or equipment. The building utilises chilled beams to cool the Studio spaces. With no moving parts, the cooling beams are extremely low noise. Attenuators were also installed in the main supply and return ductwork to the ventilation plant to prevent any noise breakout.

The nature of the occupancy requires that the studios are lobbied with acoustic doors and isolated structures within the building envelope. Sound incursion is therefore optimally reduced to a practical minimum. Three internal acoustic pods were constructed using slabs of thick precast concrete, and are separated from the floor via rubber isolating mounts to fully sound proof the area. The acoustic impact of the building occupancy on the site and surroundings is therefore also insignificant.

C. Studio Network

Using the latest Apple Mac Pro and iMac machines with Intel processors, the studios are networked using conventional gigabit ethernet. An Xserve is used for authentication and file storage, connected to the network on four gigabit ports using 802.3ad link aggregation. This server controls access to the client machines through its own Open Directory, as well as the university-wide Active Directory system. All studio computers are housed in a central hub room to eliminate mechanical noise within the studios themselves, with displays connected via optic fibre using Gefen DVI-to-Optical converters. Similarly, Firewire 800 and USB 2.0 are replicated in the studio spaces using a combination of converters over optic fibre and standard CAT5 ethernet. The audio interfaces themselves are housed in the same space as the computer systems, with bantam/PO-style patch bays providing complete flexibility over audio routing throughout the building.

D. Audio Interfaces

The primary audio interfaces are Digidesign 003 and 002 rack units, with the exception of Studio 1 which benefits from a Pro Tools HD2 system with a 192 I/O interface. All studios are also equipped with MOTU 828mkII interfaces for added flexibility. In the teaching cluster, 14 iMacs are available with a mixture of MBox 2 Pro and MBox 2 Mini interfaces. A custom switcher allows for audio to be routed to the cluster loudspeakers from any interface; a useful feature for demonstrating work in a teaching environment. All loudspeakers are manufactured by Genelec.

E. Working Spaces

1) Studio One: hosts 32-channel monitoring, and incorporates a suspended loudspeaker array on a custom made ceiling grid. This studio is mainly used for compositional, multi-channel and sound diffusion performance research. There are two sound diffusion tools: The DACS custom made MANTIS 32-channel control surface with an EtherSense device (IRCAM) and a tailor made GLUION FPGA 16bit ADC (designed by Sukandar Kartadinata) with 32 analogue inputs and 68 digital inputs/outputs for experimental purposes. A combination of MOTU 24io and 2408mkIII interfaces with Mac Pro computers serve the diffusion system.

2) Sheila Beckles Studio: dedicated for audio in multimedia production and composition, it incorporates a 5.1 Genelec monitoring system, with one Quad-core Mac Pro featuring two Dual-Core Intel Xeon processors, a 30 inch Apple Cinema Display, two additional 20 inch Cinema Displays, and other equipment. This facility is mainly designed for postgraduate and staff audio visual and compositional research.

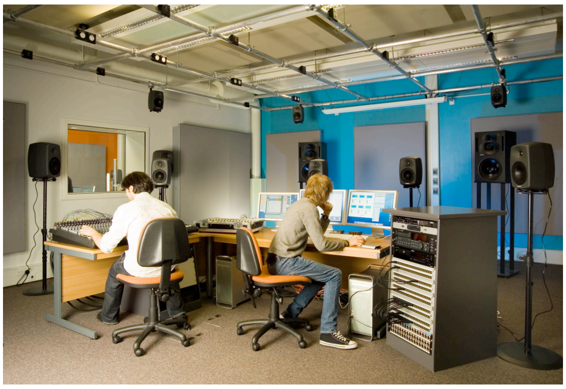


Figure 2. Postgraduate students working in Studio One

3) Studio Three: Studio Three: a 10-channel monitoring system for postgraduate and staff multi-channel compositional research. It contains 8 + 2 loudspeakers with two subwoofers (all Genelec), one Quad-core Mac Pro featuring two Dual-Core Intel Xeon processors and other equipment.

4) Studio Four: 5.1 monitoring and composition environment and also serves as a recording and teaching space, mainly for undergraduate work.

5) Studio Cluster: consists of 14 new imac workstations for group teaching and guest lectures, undergraduate and postgraduate access.

IV. MANTIS FESTIVAL

Another key aspect of NOVARS Research Centre's creative umbrella is its Music Festival called MANTIS.

Each year MANTIS (Manchester Theatre in Sound) presents several concerts and encounters in areas of electro-acoustic music and sound-art. It provides a dynamic and appealing meeting-point for composers, artists and scientists who are willing to exchange and showcase practice-led and cutting-edge computer music research. It also promotes, disseminates and performs new works from composers based at The University of Manchester.

The University's Martin Harris Centre for Music and Drama is the focal point for MANTIS concert activities, featuring a large surround sound diffusion system in the Cosmo Rodewald Hall – a 350 seat concert space, which includes box-office and support personnel who run an average of 80 concerts and events per year. Additionally, it also has other smaller venues, such as the John Thaw Studio Theatre (a black-box space specialising in drama productions) and the seminar room G16. The core of the MANTIS System consists of a 40-Genelec loudspeaker surround setup and a custom-made software and spatialisation control interface.

MANTIS has run the Festival in partnership with organisations such as Sonic Arts Network EXPO and LICA (Lancaster Institute for Contemporary Arts), both in the UK, taking the festival to numerous venues in the North of England (ie. the Victoria Baths). MANTIS has featured music from several guest composers including

Jonty Harrison, Simon Emmerson, Rajmil Fischman, Pete Stollery, Rodrigo Sigal, Francis Dhomont and Gerard Bennett. In November 2008, MANTIS will be celebrating its 10th festival in 4 years with guest composers Annette Vande Gorne (Belgium) and João Pedro Oliveira (Portugal).

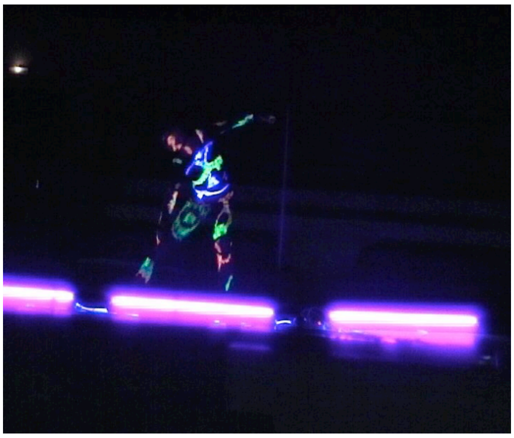


Figure 3. Idoia Zabaleta (KLEM) dancing Drosophila by Ricardo Climent at MANTIS: South-North 2007

Two very special Festival editions occurred in 2006, when MANTIS hosted guest composer Francis Dhomont for his 80th birthday and in 2007, when MANTIS celebrated the launch of NOVARS with five specially curated concerts by D. Berezan (Berezan, Bennett, Dhomont, Casken), UK (by J. Harrison), Canada (by G. Gobeil), Germany (by L. Bruemmer/F. Hein) and USA by R. Climent (with virtuoso performers Esther Lamneck and Elizabeth McNutt).

The Festival has also focused on themes such as MANTIS: South-North 07, including music from Latin American composers. The first MANTIS call for pieces received 186 submissions from all over the world. The Festival has gained the sponsorship of Arts Council England; Goethe Institute, Manchester, Sheila Beckles Foundation, Embassy of Mexico, UK; Council for Cultural Affairs, Taiwan and the Secretaria de Relaciones Exteriores in Mexico.

V.CURRENT AND FUTURE ACTIVITY

In the first year of its operation the Centre has hosted research seminars, masterclasses, workshops and fora covering electro-acoustic composition, sound diffusion technique and practice, sensors development, analysis and interdisciplinary work. Undergraduate and postgraduate course-unit teaching and work in areas of electroacoustic composition, aesthetics and acoustics have been supported, in addition to the research of 10 PhD students, 2 full-time members of academic staff and associate researchers. The KAIROS Electroacoustic Ensemble has performed in several concerts, and in 2007/08 MANTIS delivered concerts series and events in Manchester, Lancaster, Edinburgh, Brussels and Valencia. MANTIS is currently working towards a number of festival and conference events in collaboration with guest composers, artists and other institutions.

A program of Research Assemblage Groups has been established by postgraduate students and academic staff. Formulated groups focus on Sound Spatialization; Aesthetics and Analysis; Dissemination, Performance and Music Musicianship; Phonography; Synthesis and Software and Impossible Research.

NOVARS launched two Residency Schemes for Artists to undertake work in the studios (currently open to music composers) and for Engineers and physical Scientists to procure the advancement of existing or new areas of research. In 2008 NOVARS's resident artists are: João Pedro Oliveira (Portugal) sponsored by the Calouste Gulbenkian Foundation and including the commission of new work; Pippa Murphy (UK) –freelance residency- and Thomas Bjelkeborn (Sweden) sponsored by The Society of Swedish Composers and the Helge Ax:son Johnson Foundation. On the Engineering side: Sukandar Kartadinata (Germany) sponsored by the Pump Priming Fund, Manchester University, in the area of FPGAs and sensors technology; Stefan Bilbao (Canada/UK) in the area of Physical Models and Matlab and Patrick Sanan (USA) also in Physical models and Sound-art. In the next year the Centre aims to commission new creative work, continue with the organization of workshops and consolidate the efforts of the Research Assemblage Groups.

VI.ACKNOWLEDGEMENTS

Technical reports provided by Harry Fairclough Construction Ltd, Cruickshank, Seward architects ARUP consulting engineers, DACS-AUDIO and Estate Electrics; University of Manchester. Graphic materials of the building by Beccy Lane; Positive Image Photography.

For Further details visit
http://www.novars.manchester.ac.uk/

Speaker-Herd: A Multichannel Loudspeaker Project for Miscellaneous Spaces, Loudspeaker Architectures and Composition Approaches

P. Modler, L. Fütterer, E. Farchmin, F. Bierlein, R. Raepple, D. Loscher, T. Möhrmann, A. Rafinsky,
M Zielke, A. Kerschkewitz, A. Unger,
University of Media, Arts and Design, Karlsruhe, Germany

Abstract — **Strong interest in spatial sound existed throughout times and on various levels of aesthetic music and sound production and perceiption. In recent years the availability of highquality loudspeakers and digital multichannel audio systems paved the way to incorporate spatial acoustics into musical composition. In this paper we describe a project which is aimed to provide flexible possibilites to experiment with miscellaneous loudspeaker architectures and multichannel distribution systems. The system allows to use up to 96 audio channels in real time which can be fed to loudspeakers setup according to varying spatial designs. As examples a number of realised architectures and compositions will be described.**

I.Introduction

Although historic music compositions rarely approach spatial parameters as part of their composition they had been composed to be played and perceived in spaces with tangible acoustic features. E.g. medieval cathedrals support chants by adding sophisticated reverberation to provide extension and prolongation of singing voices [1]. In later periods compositions considered acoustic aspects more direct e.g. by formulating needs for acoustic reproduction spaces as well as through composition instructions considering acoustic and spatial parameters such as position of the orchestra, type of concert hall, movements of players.

The availability of loudspeakers and multichannel techniques boosted the interest to incorporate aspects of spatial listening into compositions.

In Poem Electronique precise spatial movements of acoustic events are part of the composition and were realised through loudspeaker arrays [2].

Reproduction of spatial acoustics evolved from stereo systems to surround sound systems such as Dolby DTS, Ambisonics or Wavefront-synthesis. This approaches aim to reproduce or create certain acoustic spaces such as recreating the sound image of a classical symphony recorded in a concert hall through a living-room stereo speaker setup, or more sophistaced to recreate entire acoustic wave fields through large speaker arrays [3].

A different approach follows the concept of loudspeaker orchestra (Diffusion, Acousmonium) . The aim to recreate spatial acoustic images from recorded material is waived in favour to explore characteristics of miscellaneous loudspeakers, different speaker and audience positions, different spaces and especially for the diffusion approach the interpretation by the performer who runs the mixing desk as integral part of the musical concert [4], [5].

In our approach we aim to provide a flexible environment which enables creative prototyping of experimental loudspeaker setups which are not bound to certain rooms or spaces and are open to applications of all kinds of multichannel sound distribution strategies and software. Instead of relying to large scale and so inflexible and fixed setups or built in speaker systems in concert halls we aim to provide tools to setup speaker and mutlichannel software for miscellanous architectures and all kind of buildings or spaces.

II.Hardware Components

Based on the aim to realise a flexible lowcost system but at the same time provide as far as possible high quality multichannel audio we decided to choose five basic components:

- A large but still extendable number of hig-quality loudspeakers
- A multichannel digital audio interface with at least 64 audio channels
- A powerfull cpu computer for multichannel mixing and spatialisation algorithms
- Necessary cabeling e.g. from the audio-interface to the speakers
- Rigging or stands for the speakers or additional gear for certain speaker setups

A. Speakers

For the loudspeakers we choose Genelec 8040 and 8240 which was based upon following considerations:

- A flat free field frequency resonse of a 48 Hz - 20 kHz (± 2 dB) at at 105 dB SPL@1m
- Compact physical dimensions of 350 x 237 x 223 mm and 8.6 kg weight
- Extremly low channel crosstalk
- High audio quality of amplifier and speaker

We accepted the drawback of additional cabeling for the power supply of the speakers in favour of the high audio reproduction features and the compact dimensions of the speakers. [6]

B. Processing Computer

Based on the aim to provide the main operating system platforms used for audio production an Apple Intel machine (MacPro) was choosen with a quad-core kernel. The machine can be booted into OS-X, Windows and Linux, and is equipped with a second grafics card enabling to run up to 4 separate screens.

C. *Digital Multichannel Audio Interface*

Due to budget constraints to keep the costs of the project in certain range we were searching for a solution poffering a large number of high quality audio channels at a low cost per channel rate. Due to the ePCI bus of the MacPro a range of audio interfaces were possible.

A Digidesign/Protools solution would have provided highest audio quality but at a high price per audio channel.

RMEs MADI solution gives 64 audio channel with one PCI card and the possibility to synchronise a second card with additional 64 channels as a tandem solution. The RME-MADI system provides easy distibution of audio through digital MADI conenctions (BNC) or after a conversion through ADAT or AES/EBU connections. In combination with digital speakers this enables flexible and comfortable cabling betweent the audio interface and the speakers, also in large and complex setups.

A third alternative is MOTUs 444 ePCI card combined with up to four 24IO converter units giving a total of 96 input and 96 output channels. In contrast to RMEs MADI system the 24IO box is a straight digital to analog and analog to digital converter box with 24 balanced analog inputs and outputs. An interesting feature of the 444 card is the possibility to connect most of MOTUs audio interfaces, including older 2408 interfaces, which provide up to 24 ADAT I/Os. With this option it is possible to connect digital drains or sources to the 444 in combination with 24IO units.

In short the RME-MADI is a fully digital system, providing a large total number of audio channels (128 in tandem mode) at a reasonable low cost per channel rate using digital speakers. The MOTU solution has even a smaller cost per channel rate with the tradeoff of a maximim number of 96 channels. Due to the analog I/Os of the 24IO more effort has to be put into cabling e.g. multicores for larger distances.

Based on our aim to use only a certain budget in combination with the possibility to use existing material, such as multicores, balanced leads, power supply units, we choose the MOTU 444 ePCI card with 24IO analog interfaces and a 2408 digital interface which was already at hand.

III. SOFTWARE

A. *Operating Systems*

Due to the versatile multi boot option of the Intel MacPro machine it is possible to work on the most commonly used operating systems: OSX, Windows, Linux.

B. *Software for Audio, Spatialisation and Distribution*

Most of the compositions were realised on Max/Msp/Jitter on OSX. Pieces which had been projected in Max/Msp on Windows were converted to OSX, which enabled a faster loading of different setups. A smaller number pieces were realised in PD running on Windows.

For the spatialisation and sound distribution following software packages are accessible:

- ICST Ambisonics Spatialisation Tools [7]
- SPAT IRCAM [8]
- vbab~ [9]
- Nuendo/Cubase [10]
- Zirkonium [11]

Besides this spatialisation tools a number of pieces were realised with non-standard audio distribution approaches which had been programmed and realised by the composers.

IV. SPEAKER ARCHITECTURES / SPEAKER SCULPTURES

All setups and compositions are in mutlichannel format, with a current number of channels of about 40-50 channels.

A. *Dome*

A Dome setup e.g. as a geodatic dome.

B. *Stereo Line or Pathway:*

This is a pathway through speakers setup at ear level.
The distance between the speakers is about 1.5m. Audience are tracked through a camera/computer system which may be used to control the sound compositions. In the current setup the pathway is about 30m length.

C. *Cone or Tree*

Loudspeakers suspended from a rigging are arranged in form of a cone. The cabeling of the speakers gives the impression of root, trunk and branches of a tree.

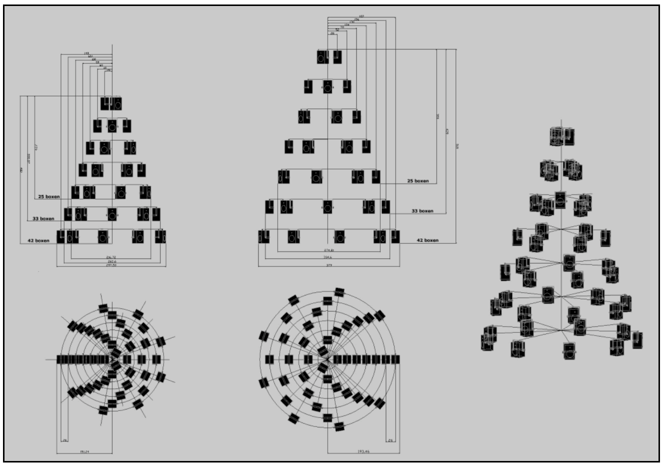


Fig. 1: 2 cone/tree setups with 42 Speakers, ~6m height and ~5m base diameter (CAD)

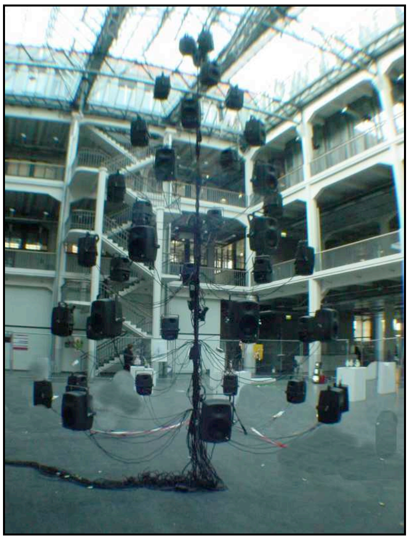


Fig. 2: sOUNDtREE: Cone/Tree setup

D. Full spheric setup: Sound-Sphere

A full sphere setup with speakers. Ear level of the auditors is about the centre of the sphere. Auditors access the sphere through a stairway onto a platform of about 3x3 sqm size and 1m height. Various sound spatialisation approaches are explored through different compositions.

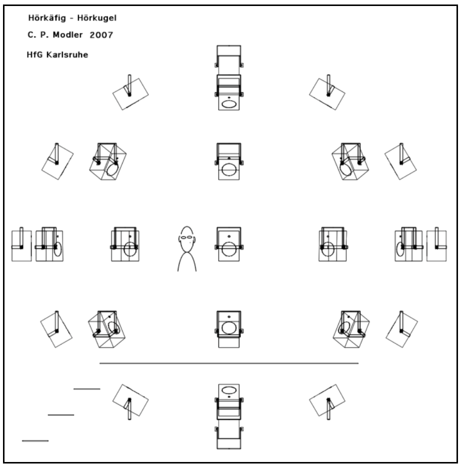


Fig. 3: Full Sphere: Frontview (CAD)

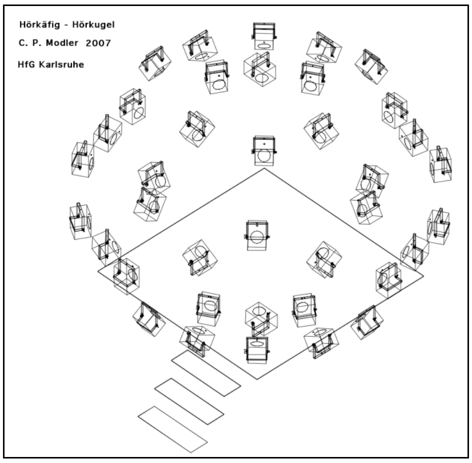


Fig. 4: Full Sphere: 3D-view (CAD)

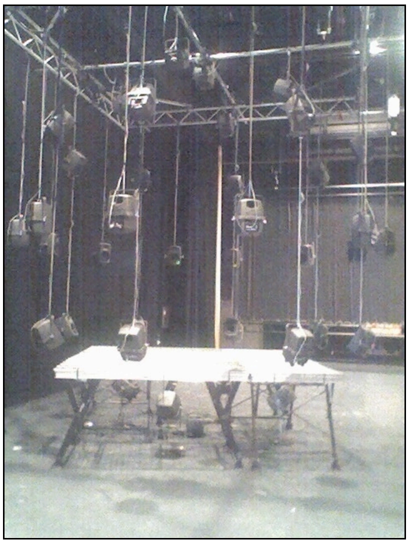


Fig. 5: sOUNDsPHERE: Full Sphere setup

E. Mixed

A mixed setup e.g. a setup designed by T. Myatt for his piece untitled-3 which was premiered on this sytem. The architecture was based on a geodatic dome with 13 speakers, a supporting inner circle close to the audience and two larger rectangles about a sice of 18x24m with each 8 speakers. One rectangle at floor level and the second rectangle in about 7m height.

F. Studio

A setup to provide flexible setup in a standard studio space. The speakers are supended with standard theater hooks and can be arranged freely.

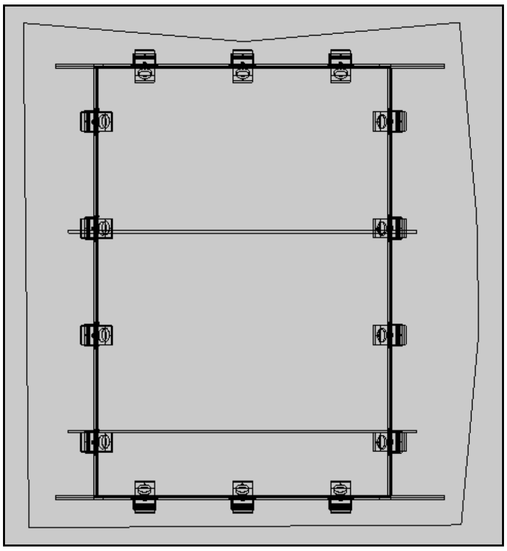


Fig. 6: Rectangular studio setup topview (CAD)

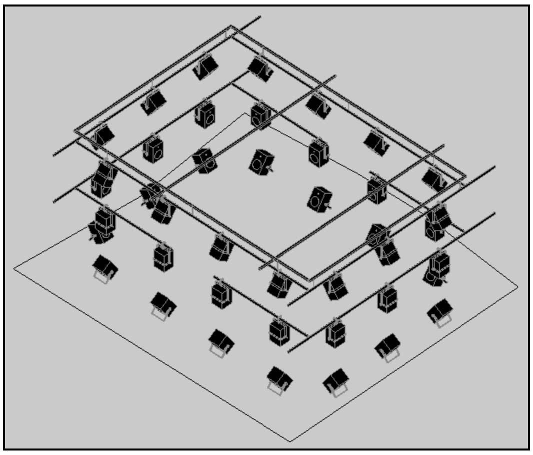


Fig. 7: Rectangular studio setup 3d-view (CAD)

V. SPATIALISATION AND DIFFUSION APPROACHES

In this section a number of approaches are described, which had been used for the spatialisation of audio in the above described loduspeaker setups. The approaches reach form vector panning and ambisonics to non-standard algorithms which had been implemented for the individual piece. In the following a number of compositions are shorlty described.

A. Full-Sphere: Fetzenfische (L. Fuetterer)

Fetzenfische is an algorithmic composition realised with reaktor (native instruments) and Max/Msp. The spatialisation for 42 audio outputs is based on two identical algorithms. The signal of one algorhtm runs over 2 channels (a,b). Each position of a speaker in the sphere owns a matrix giving the adjacent positions. After a complete movmenent from a to b is accomplised the source moves on according to the matrix values.

B. Cone/Tree: 42 Channel Plunderphonic Christmas Song (M. Zielke, D. Loscher)

This Song is made for 42 speakers built like a Christmas Tree. From 42 famous Christmas Songs there were only taken the relevant phrases: christmas, tree, a child is born, Jesus, holy night, star, Bethlehem ... .The result sounds like a holy cacophony.

C. Stereo Line/Pathway: Karlsruhe klingt (D. Loscher)

Every city has its own soundscape: By walking through the line of 42 Speakers (21 on each side) you hear 21 field recordings characterizing the city of Karlsruhe. At the beginning of the row there´s somebody asking how the city sounds for you: By walking forward the runaway will notice that there are skaters in the park, trams, cars, firecracker from New Year´s eve, passersby, children playing, sine waves from buildings & current generators, crowds, shopping malls, leaves, birds, stairs climbing...

VI.Results

The speaker herd project provides the possibility to experiment with a larger number of similar loudspeakers (currently 42 units). The audio distribution system based on a MacPro quad kernel machine combined with a MOTU 444 ePCI, two 24IO units and a 2408 unit offers sufficient CPU power for running spatialisation tools in realtime and to quite a large extent composition and sound generation tasks.

For the realisation of a certain speaker architecture quite a large amount of time was spent to plan the physical setup of the speakers such as precise riggin, suspension etc. The use of a CAD system supported the design process, both in providing fast access to measures but also to have a visuall feedback of the projected speaker setup.

A tedious part of each setup was the cabling. It turned out that indexing cables and boxes were not necessary. Instead most times it was faster to connect speakers and cables with the audio interface at random, and later sort the connections through a software layer. This was easy to achieve by sequentially detecting connections of single speakers through sending noise to each audio output, and then adjusting the indexing of the outputs. E.g. in Max/Msp this can be achieved through the I/O Mappings page.

Line level adjustment of the outputs of the MOTU 24IO interface in combination with the Genelc 8040A speakers turned out to be difficult, since the 24IO outputs do not provide a level reduction. For most cases the input level of the 8040s had to be reduced to a large extent due to the high level the MOTU interface was delivering.

The pieces so far realised on the speaker herd may be categorised into pieces which try to approach audio distribution through a source/space model (e.g. ambisonics) or a non-standard diffusion approach (e.g. randomo distributions) or a combination of both.

Quite a number of pieces tried to leave the source/space model and define their own way to distribute audio onto a large number of speakers. E.g. F. Bierleins piece followed the paradigm of a granulated spoken sentence where each grain of the sentence is sent to a certain speaker at a certain time. The auditor may then perceive seqeuntially the whole sentence by walking through the whole line of speakers. L. Fuetteres piece for the sound sphere does also not use a standard spatialisation approach but instead distributes grains of saound material to speakers depending on algorithmic computation. In contrast to this approaches the swarm uses the ICST tools for ambisonic spatialisation with quite a large numebr of virtual audio sources in space and controlled random movements to create the imrpession of a swarm of flying animals.

Although a detailed evaluation of the spatialisation software is beyond this paper following aspects were realsied during the project. The ICST ambisonics Max/Msp tools provide a sophisticated and elaborate interface for ambisonics in Max/Msp. Although the tools were used in several compositions it is far away from having explored all possiibilities. An interesting point in the ambisonics realsiation was that height worked in a straigt plane setup well but not in a speaker setup distributed over three diemensions. We experienced this with a classical doem setup and with the full spehere setup where in both setups the height of audio sources especially in below the ear were difficult to reproduce.

VII.Conclusions

The speaker herd provided an exceptional possibility to experiment with loudspeaker architectures and sound distribution and spatialisation algorithms. The choosen hardware and software configuration which is based on a MacPro quad kernel CPU, a MOTU 96 channel digital audio interface and a larger number of Genelec 8040s and 8240s provided a robust and extendable basis for various speaker architectures, audio distribution approaches and compositions. Although for certain compositions and situations speakers with higher power output may be desirable, the 8040 speaker provided extremly high audio quality combined with compact dimensions which enabled flexible and non standard speaker setups.

The full sphere setup explores sound sources from below ear level which is not reflected in most of the standard approaches such as e.g. domes or wavefield synthesis. The stereo pathway setup animates auditors to move along a lin of speakers, wheras the cone/tree setup the explores the acoustics of the installation space due to its circular acoustic radiation.

Quite a number of further speaker architectures are planned to be realised based on this sytem and a collaborative group of artists, musicians and composers.

Acknowledgments

This work was realised with the support of a grant from the Ministery of Education and Research of Baden-Wuerttemberg, Germany. Thanks to T. Myatt, University of York, UK and T. Troge, Hochschule für Musik, Karlsruhe for their support.

References

[1] B. Shield, T., Cox, Concert Hall Acoustics: Art and Science, http://www.acoustics.salford.ac.uk/acoustics_info/concert_hall_acoustics/?content=musical_acoustics, 1999/2000

[2] http://www.music.columbia.edu/masterpieces/notes/varese/notes.html, 2008

[3] R. Pellegrini, M. Rosenthal, and C. Kuhn, "Wave field synthesis: open system architecture using distributed processing," in Forum Acusticum, Budapest, Hungary, September 2005.

[4] Austin, Larry (2001). "Sound Diffusion in Composition and Performance Practice II: An Interview with Ambrose Field". Computer Music Journal 25 (4): 24. MIT Press.

[5] Harrison, Jonty (1988), "Space and the BEAST concert diffusion system", written at Ohain, in Francis Dhomont, L'espace du son, Musiques et Recherches

[6] http://www.genelec.com/technical-documents/, 2008

[7] Jan C. Schacher, Philippe Kocher, Ambisonics Spatialization Tools for Max/MSP, Proceedings of the 2006 International Computer Music Conference, New Orleans, 2006

[8] Jot, J.M., Warusfel, O. 1995a "Spat: a spatial processor for musicians and sound engineers", Proc. CIARM'95 Conference, Ferrara (Italie), Mai 1995.

[9] Pulkki, V. Creating generic soundscapes in multichannel panning in Csound synthesis software. Org. Sound 3, 2 (Aug. 1998), 129-134.

[10] http://www.steinberg.net/, 2008

[11] http://on1.zkm.de/zkm/stories/storyReader$5970

Modeling Affective Content of Music: A Knowledge Base Approach

António Pedro Oliveira*, Amílcar Cardoso*
*Centre for Informatics and Systems of the University of Coimbra, Coimbra, Portugal

***Abstract*—The work described in this paper is part of a project that aims to implement and assess a computer system that can control the affective content of the music output, in such a way that it may express an intended emotion. In this system, music selection and transformation are done with the help of a knowledge base with weighted mappings between continuous affective dimensions (valence and arousal) and music features (e.g., rhythm and melody) grounded on results from works of Music Psychology.**

The system starts by making a segmentation of MIDI music to obtain pieces that may express only one kind of emotion. Then, feature extraction algorithms are applied to label these pieces with music metadata (e.g., rhythm and melody). The mappings of the knowledge base are used to label music with affective metadata. This paper focus on the refinement of the knowledge base (subsets of features and their weights) according to the prediction results of listeners' affective answers.

I. Introduction

Music has been widely accepted as one of the languages of emotional expression. The possibility to select music with an appropriate affective content can be helpful to adapt music to our affective interest. However, only recently scientists have tried to quantify and explain how music expresses certain emotions. As a result of this, mappings are being established between affective dimensions and music features [14][8].

Our work intends to design a system that may select music with appropriate affective content by taking into account a knowledge base with mappings of that kind. Most psychology researchers agree that affect has at least two distinct qualities [13][16][3]: valence (degree of satisfaction) and arousal (degree of activation), so we are considering these 2 dimensions in the classification. Automated classification using machine learning approaches has the advantage of allowing one to perform classifications in a faster and more reliable way than manual classifications. We intend to improve the knowledge base by selecting prominent features and by defining appropriate weights. This is done, respectively, by using feature selection and linear regression algorithms.

The automatic selection of music according to an affective description has a great application potential, namely in entertainment and healthcare. On the one hand, this system can be used in the selection of soundtracks for movies, arts, dance, theater, virtual environments, computer games and other entertainment activities. On the other hand, it can be used in music therapy to promote an intrinsic well-being. The next section makes a review of some of the most relevant contributions from Music Psychology and related works from Music Information Retrieval. Section III gives an overview of the system. Section IV presents the details of the experiment. Section V shows the experimental results. Section VI analyses the results, and finally, section VII makes some final remarks.

II. Related work

This work entails an interdisciplinary research involving Music Psychology and Music Information Retrieval. This section makes a review of some of the most relevant contributions for our work from these areas.

A. Music Psychology

Schubert [14] studied relations between emotions and musical features (melodic pitch, tempo, loudness, texture and timbral sharpness) using a 2 Dimensional Emotion Space. This study was focused on how to measure emotions expressed by music and what musical features have an effect on arousal and valence of emotions. Likewise, Korhonen [4] tried to model people perception of emotion in music. Models to estimate emotional appraisals to musical stimuli were reviewed [14][6] and system identification techniques were applied. Livingstone and Brown [8] provided a summary of relations between music features and emotions, in a 2 Dimensional Space, based on some research works of Music Psychology. Gabrielsson and Lindstrom [3] is one of these works, where relations between happiness and sadness, and musical features are established. Lindstrom [7] analysed the importance of some musical features (essentially melody, but also rhythm and harmony) in the expression of appropriate emotions.

B. Music Information Retrieval

Emotions detection in music can be seen as a classification problem, so the selection of the classifier model and the feature set are crucial to obtain good results. Van de Laar [17] compared 6 emotion detection methods in audio music based on acoustical feature analysis. Four central criteria were used in this comparison: precision, granularity, diversity and selection. Emotional expressions can be extracted from music audio [19]. The method designed by Wu and Jeng consisted in 3 steps: subject responses, data processing and segments extraction. From the results of this method, emotional content could be associated to musical fragments, according to some musical features like pitch, tempo and mode.

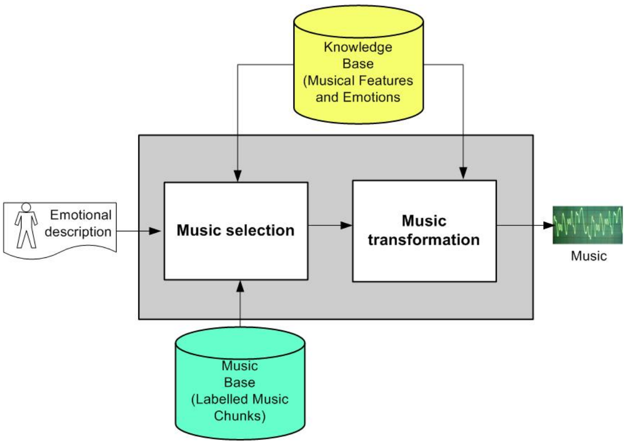

Knowledge Base (Musical Features and Emotions
Emotional description
Music selection
Music transformation
Music
Music Base (Labelled Music Chunks)

Fig. 1. System overview

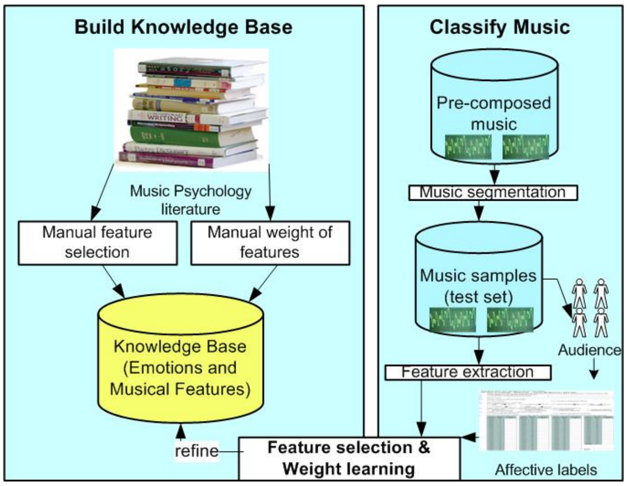


Fig. 2. Stages of the experiment

Muyuan and Naiyao [10] made an emotion recognition system to extract musical features from MIDI music. Support Vector Machines were used to classify music in 6 types of emotions (e.g., joyous and sober). Both statistical (e.g., pitch, interval and note density) and perceptual (e.g., tonality) features were extracted from the musical clips. There are also models to recommend MIDI music based on emotions [5]. The model of Kuo et al., based on association discovery from film music, proposes prominent musical features according to detected emotions. These features are compared with features extracted from a music database (chord, rhythm and tempo). Then, the result of these comparisons is used to rank music and a list of recommended music is given according to 15 groups of emotions.

III. Project description

The work described in this paper is part of a project that has the objective of implementing and assessing a computer system that can control the affective content of the music output, in such a way that it may express an intended emotion. The system uses a database of pre-composed music represented at a symbolic level. We intend to accomplish our objective in 2 stages. The first consists in the selection / classification of music by affective content and is the focus of this paper. The second stage will deal with the transformation of the selected music to approximate even further its affective content to an intended emotional description. These stages are done with the help of a knowledge base with weighted mappings between continuous affective dimensions (valence and arousal) and music features (e.g., rhythm and melody). Fig. 1 illustrates our system.

IV. Details of the experiment

The experiment here described follows a preliminary one [12], overcoming some of its limitations by using larger numbers of music files, listeners and music features. Fig. 2 presents an overview of different stages of our experiment. The process starts with the segmentation of MIDI music to obtain segments that may express only one kind of emotion (this method is described in detail in the following paragraph). Then, feature extraction algorithms of third party software [9][2][15] are applied to label these segments with music metadata (e.g., rhythm and melody). The mappings of the knowledge base are used to label music with affective metadata. For the experiments with listeners, we used a test set of 96 musical pieces. These pieces were of western tonal music (film music), last, approximately, from 20 seconds to 1 minute, and were used for both training and validating the classifier. 80 different listeners were asked to label online each affective dimension of the musical pieces with values selected from the integer interval between 0 and 10 [0;10][1]. The obtained affective labels were used to refine the sets of features and corresponding weights in the knowledge base. This was done separately for the valence and arousal.

A. *Music segmentation*

The expression of emotions in music varies as a function of time [4]. To facilitate classification, it is very important to obtain segments of music that may express only one kind of affective content. Our segmentation module uses the Local Boundary Detection Model (LBDM) [1][2] to obtain weights based on the strength of music variations (pitch, rhythm and silence). These weights establish plausible points of segmentation between segments with different musical features that may reflect different affective content. We define a threshold to reduce the search space among the LBDM weights. This threshold is equal to 1.30*mean(LBDM weights)+1.30*standard deviation(LBDM weights). We obtain music chunks of different length with a minimum of notes *MinN* and a maximum of notes *MaxN*. To segment, we start at the beginning of the MIDI file and look for a plausible point of segmentation that corresponds to the maximum weight between the beginning of MIDI file+*MinN* and the beginning of MIDI file+*MaxN*. This process is repeated starting from the last point of segmentation until we come to the end of the MIDI file.

[1] http://student.dei.uc.pt/%7Eapsimoes/PhD/Music/smc08/index.html

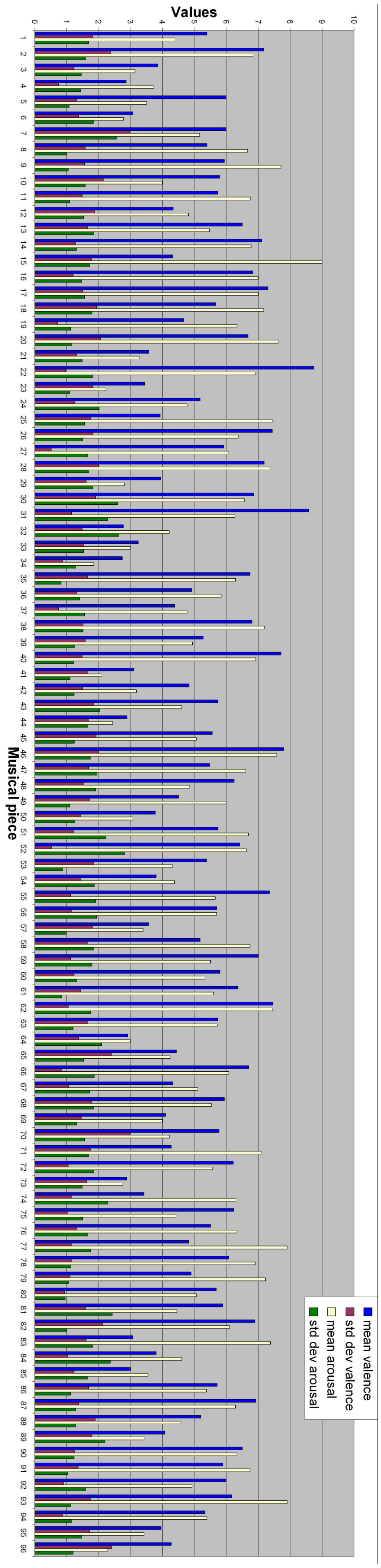


Fig. 3. Mean and standard deviations of the affective responses for valence and arousal

B. Selection and weighting of features

We examined 146 unidimensional features and 3 multidimensional ones that were categorized in 6 groups: instrumentation (20), texture (15), rhythm (39), dynamics (4), melody (68) and harmony (3). Special attention was devoted to new features and important ones from a preliminary experiment[12]: the importance (volume*time) of 13 MFCCs [17] of each sample used to synthesize musical instruments, the prevalence (by note or time) of specific groups and individual instruments, tempo, notes density, duration of notes, rhythmic variability, melodic complexity, number of repeated notes, prevalence of the most common melodic intervals, pitch classes and pitches, and mode (major or minor). Each feature was analysed for the affective dimensions with the help of the affective labels obtained for the test set and of information obtained from the literature [11]. This was done by applying the following feature selection algorithms: Genetic search, best-first and greedy stepwise [18].

With the subsets of features selected, some algorithms of linear regression were used to refine the weights of each feature. Linear regression, SMO regression and SVM regression [18] were tested. The results of the next section were obtained with SVM regression, because it was, generally, the approach that gave us the best results.

V. Results

Fig. 3 shows the mean and standard deviation for affective responses obtained in the online questionnaire[2]. Answers distant more than the mean ± 2*standard deviation were discarded.

The importance of individual features in each group of features was established (represented as positive or negative between parenthesis in tables I and II). All the features presented in tables I and II have a correlation coefficient higher than 15% with the affective labels.

A. Valence

Table I presents prediction results by groups of features for valence. From this, we can infer that rhythmic (e.g, tempo, average note duration, variability of note duration and time between onsets), harmonic (e.g., key mode and key), melodic (e.g., climax position and melodic complexity), texture (e.g., spectral texture MFCC 4 and 6, and number of unpitched instruments) and instrumentation features (e.g., string ensemble fraction) are relevant to the valence of music.

We started by applying feature selection algorithms [18] to reduce the number of features and to improve classification results. From this a group of 26 features resulted. The correlation and determination coefficients for training on the whole set were, respectively, 89.37% and 79.86%. 8-fold cross validation of classification resulted in correlation and determination coefficients of, respectively, 81.21% and 65.95%. After this we selected manually the best group of features to know the most important features in the stage of selection, but also for the stage of transformation. From this a group of 5 features

Features	Cor. Coef.	Det. Coef.
Note Prevalence Muted Guitar (+)	36.99%	13.68%
Electric Instrument Fraction (+)	33.72%	11.37%
Note Prevalence Steel Drums (+)	33.21%	11.02%
Time Prevalence Marimba (+)	31.41%	9.86%
Note Prevalence Fretless Bass (+)	31.02%	9.62%
Note Prevalence Timpani (-)	26.76%	7.16%
Electric Guitar Fraction (+)	23.4%	5.47%
String Ensemble Fraction (-)	21.5%	4.62%
Note Prevalence Pizzicato Strings (-)	21.08%	4.44%
Orchestral Strings Fraction (-)	20.7%	4.28%
Note Prevalence Orchestral Harp (-)	20.37%	4.14%
Saxophone Fraction (+)	19.75%	3.9%
Note Prevalence English Horn (-)	19.69%	3.87%
Note Prevalence French Horn (-)	19.56%	3.82%
Note Prevalence Tenor Sax (+)	19.18%	3.68%
Note Prevalence Synth Brass 1 (+)	19.12%	3.65%
Note Prevalence Pad 3 (polysynth) (+)	18.66%	3.48%
Note Prevalence Bassoon (-)	18.49%	3.41%
Time Prevale. Acoustic Grand Piano (-)	16.71%	2.79%
Acoustic Guitar Fraction (+)	16.46%	2.71%
Note Prevalence Ocarina (-)	16.18%	2.62%
Note Prevalence Banjo (-)	16.18%	2.62%
Note Prevalence Flute (-)	16.16%	2.61%
Woodwinds Fraction (-)	16.12%	2.60%
Note Prevalence Tuba (-)	15.88%	2.52%
Note Prevalence Xylophone (+)	15.0%	2.25%
Note Prevalence Accordion (+)	15.0%	2.25%
Spectral Texture MFCC 4 (+)	22.89%	5.23%
Spectral Texture MFCC 6 (+)	22.45%	5.04%
Spectral Texture MFCC 7 (+)	20.85%	4.35%
Number of Unpitched Instruments (+)	20.27%	4.11%
Spectral Texture MFCC 8 (+)	17.64%	3.11%
Spectral Texture MFCC 12 (-)	17.14%	2.94%
Number of Pitched Instruments (+)	16.39%	2.69%
Relative Note Density of Highest Line (-)	15.55%	2.42%
Initial Tempo (+)	62.95%	39.63%
Average Note Duration (-)	49.92%	24.92%
Average Time Between Attacks (-)	48.72%	23.73%
Strength Strong. Rhythmic Pulse (-)	42.72%	18.25%
Variability of Note Duration (-)	42.41%	17.98%
Note Density (+)	40.99%	16.8%
Strength Two Strong. Rhythmic Pulses (-)	37.66%	14.18%
Variability of Time Between Attacks (-)	36.57%	13.37%
Number of Relatively Strong Pulses (+)	30.24%	9.14%
Distinct Rhythm Count (+)	29.03%	8.43%
Rhythmic Variability (-)	28.06%	7.87%
Strength Sec. Strong. Rhythmic Pulse (-)	25.58%	6.54%
Strongest Rhythmic Pulse (+)	20.71%	4.29%
Average Meter Accent Synchrony (+)	19.88%	3.95%
Polyrhythms (-)	18.66%	3.48%
Staccato Incidence (+)	15.05%	2.26%
Climax Position (+)	32.7%	10.69%
Average Melodic Complexity (+)	24.15%	5.83%
Interval Strong. Pitch Classes (+)	20.84%	4.34%
Dominant Spread (+)	20.83%	4.34%
Consecutive Identical Pitches (+)	18.42%	3.39%
Key mode (-)	43.86%	19.23%
Key (-)	37.79%	14.28%
Strong Tonal Centres (-)	17.43%	3.04%

TABLE I
BEST FEATURES OF EACH GROUP - VALENCE

Features	Cor. Coef.	Det. Coef.
Electric Instrument Fraction (+)	28.48%	8.11%
String Ensemble Fraction (-)	27.79%	7.72%
Note Prevalence English Horn (-)	26.15%	6.84%
Number of Unpitched Instruments (+)	25.56%	6.53%
Note Prevalence Flute (-)	25.09%	6.29%
Brass Fraction (+)	25.0%	6.25%
Note Prevalence Orchestra Hit (+)	22.97%	5.28%
Electric Guitar Fraction (+)	21.5%	4.62%
Woodwinds Fraction (-)	21.08%	4.44%
Saxophone Fraction (+)	20.78%	4.32%
Percussion Prevalence (+)	20.75%	4.30%
Note Prevalence Tremolo Strings (+)	19.52%	3.81%
Note Prevalence Orchestral Harp (-)	18.96%	3.59%
Note Prevalence Electric Bass (finger) (+)	18.9%	3.57%
Time Prevalence Acoustic Guitar (nylon) (-)	17.79%	3.16%
Spectral Texture MFCC 2 (+)	28.16%	7.93%
Variab. Prevalence Unpitched Instruments (+)	25.86%	6.69%
Spectral Texture MFCC 4 (+)	24.82%	6.16%
Melodic Intervals in Lowest Line (-)	18.99%	3.61%
Relative Range of Loudest Voice (-)	17.84%	3.18%
Average Note Duration (-)	68.67%	47.15%
Note Density (+)	63.59%	40.44%
Variability of Note Duration (-)	57.4%	32.94%
Initial Tempo (+)	55.52%	30.82%
Average Time Between Attacks (-)	55.32%	30.6%
Variability of Time Between Attacks (-)	54.07%	29.23%
Average Duration Accent (-)	53.81%	28.95%
Strength Strongest Rhythmic Pulse (-)	47.58%	22.64%
Number of Relatively Strong Pulses (+)	43.86%	19.24%
Strength Two Strong. Rhythmic Pulses (-)	41.69%	17.38%
Polyrhythms (-)	38.33%	14.69%
Strongest Rhythmic Pulse (+)	35.51%	12.61%
Strength Second Strong. Rhythmic Pulse (-)	27.9%	7.78%
Onset Autocorrelation (-)	26.67%	7.11%
Syncopation (-)	25.36%	6.43%
Average Meter Accent Synchrony (-)	24.15%	5.83%
Number of Strong Pulses (+)	23.64%	5.59%
Rhythm Range (-)	23.46%	5.50%
Rhythmic Variability (-)	23.02%	5.30%
Staccato Incidence (+)	35.22%	12.40%
Average Range of Glissandos (-)	17.51%	3.07%
Climax Position (+)	45.39%	20.60%
Average Melodic Complexity (+)	38.4%	14.74%
Consecutive Identical Pitches (+)	37.06%	13.73%
Climax Strength (-)	33.12%	10.97%
Repeated Notes (+)	32.85%	10.79%
Most Common Pitch Class Prevalence (+)	31.59%	9.98%
Relative Strength of Top Pitch Classes (-)	30.69%	9.42%
Amount of Arpeggiation (+)	29.74%	8.84%
Same Direction Interval (+)	27.95%	7.81%
Repeated Pitch Density (+)	24.46%	5.98%
Most Common Pitch Prevalence (+)	24.39%	5.95%
Distance Common Melodic Intervals (+)	22.61%	5.11%
Overall Pitch Direction (+)	21.78%	4.74%
Most Common Melodic Interval Prevale. (+)	21.11%	4.46%
Melodic Octaves (+)	20.32%	4.13%
Melodic Thirds (-)	18.04%	3.25%
Interval Between Strongest Pitch Classes (+)	17.89%	3.2%
Duration of Melodic Arcs (+)	17.67%	3.12%
Key mode (-)	22.13%	4.90%

TABLE II
BEST FEATURES OF EACH GROUP - AROUSAL

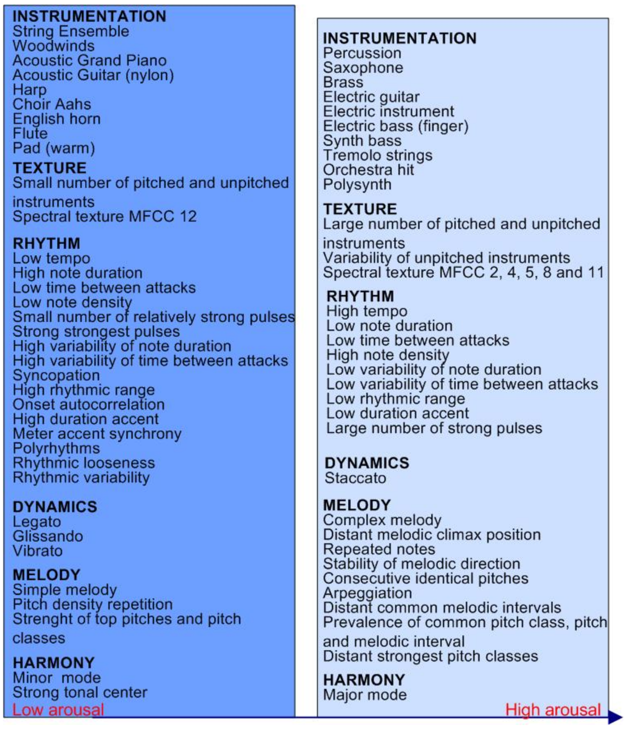


Fig. 4. Mappings for arousal

resulted. The correlation and determination coefficients for training on the whole set were, respectively, 74.95% and 56.17%. 8-fold cross validation of classification resulted in correlation and determination coefficients of, respectively, 71.5% and 51.12%. Valence is calculated by the weighted sum of the best features: -0.41*average note duration + 0.17*dominant spread + 0.41*initial tempo - 0.18*key mode + 0.24*climax position.

Correlation and determination coefficients of 56.5% and 31.92% exists between the affective labels and results obtained using the weighted mappings of a preliminary experiment[1].

B. Arousal

Table II presents prediction results by groups of features for arousal. From this, we can infer that rhythmic (e.g., average note duration, note density, time between attacks and variability of note duration), dynamics (e.g., staccato incidence), texture (spectral texture MFCC 4 and strength of top pitch classes), melodic (e.g., climax position and repeated notes) and instrumentation features (e.g., number of unpitched instruments and brass fraction) are relevant to the arousal of music.

We started by applying feature selection algorithms [18] to reduce the number of features and to improve classification results. From this a group of 23 features resulted. The correlation and determination coefficients for training on the whole set were, respectively, 90.31% and 81.55%. 8-fold cross validation of classification resulted in correlation and determination coefficients of, respectively, 84.14% and 70.79%. After this we manually selected the best group of features to know the most important features in the stage of selection, but also for the stage of transformation. From this a group of 4 features resulted. The correlation and determination coefficients for training on the whole set were, respectively, 83.86% and 70.32%. 8-fold cross validation of classification resulted in correlation and determination coefficients of, respectively, 79.14% and 62.63%. Arousal is calculated by the weighted sum of the best features: -0.56*average note duration + 0.24*initial tempo + 0.11*climax position + 0.37*consecutive identical pitches + 0.58*note density.

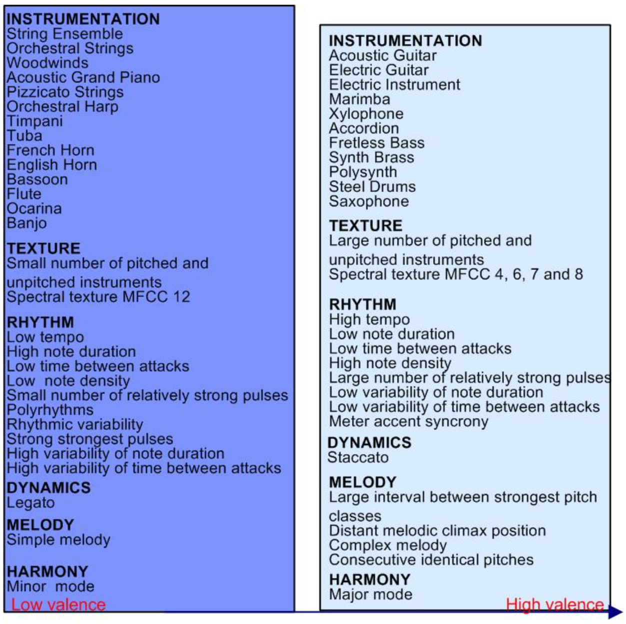


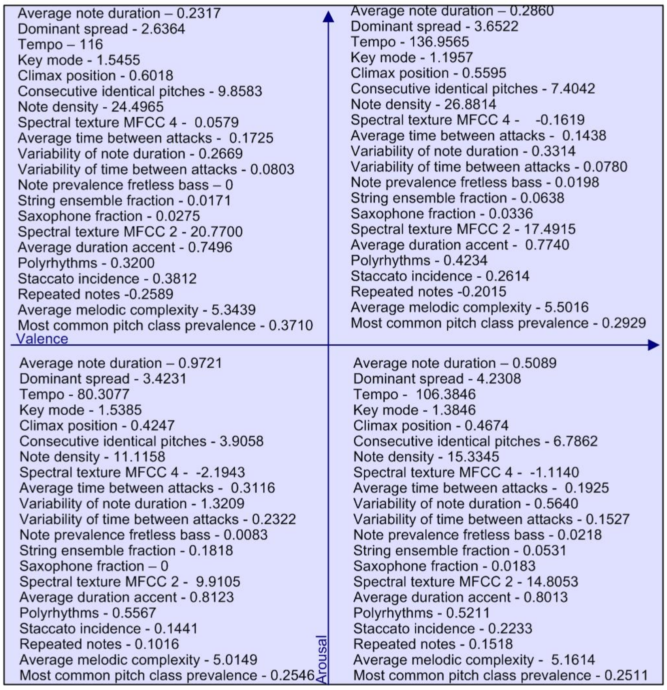
Fig. 5. Mappings for valence

Correlation and determination coefficients of 76.58% and 58.64% exists between the affective labels and results obtained using the weighted mappings of a preliminary experiment[1].

VI. Discussion

From this work much information was obtained. Fig. 4 and 5 illustrate how specific types of musical features can be changed to shift, respectively, the arousal and valence of music. For instance, a decrease in the duration of notes contribute to a decrease of both arousal and valence. Fig. 6 presents the mean of important features for each quadrant (e.g., high valence and low arousal). All this information is stored in the knowledge base and will be used in the next stage of our work that deals with the transformation of music affective content (Fig. 1).

With similar goals to [5] [10], we have developed a regression model (knowledge base) with relations between music features and emotions, Kuo et al. developed an affinity graph and Muyuan and Naiyao a SVM classifier. We used continuous dimensions (valence and arousal) instead of discrete emotions ([5] [10]). The results of our model (≈ 90%) surpass the results of Kuo et al. (≈ 80%) and Muyuan and Naiyao for valence (≈ 70%) when using a higher number of features (≈ 20).



Average note duration – 0.2317
Dominant spread - 2.6364
Tempo – 116
Key mode - 1.5455
Climax position - 0.6018
Consecutive identical pitches - 9.8583
Note density - 24.4965
Spectral texture MFCC 4 - 0.0579
Average time between attacks - 0.1725
Variability of note duration - 0.2669
Variability of time between attacks - 0.0803
Note prevalence fretless bass – 0
String ensemble fraction - 0.0171
Saxophone fraction - 0.0275
Spectral texture MFCC 2 - 20.7700
Average duration accent - 0.7496
Polyrhythms - 0.3200
Staccato incidence - 0.3812
Repeated notes -0.2589
Average melodic complexity - 5.3439
Most common pitch class prevalence - 0.3710

Average note duration – 0.2860
Dominant spread - 3.6522
Tempo - 136.9565
Key mode - 1.1957
Climax position - 0.5595
Consecutive identical pitches - 7.4042
Note density - 26.8814
Spectral texture MFCC 4 - -0.1619
Average time between attacks - 0.1438
Variability of note duration - 0.3314
Variability of time between attacks - 0.0780
Note prevalence fretless bass - 0.0198
String ensemble fraction - 0.0638
Saxophone fraction - 0.0336
Spectral texture MFCC 2 - 17.4915
Average duration accent - 0.7740
Polyrhythms - 0.4234
Staccato incidence - 0.2614
Repeated notes -0.2015
Average melodic complexity - 5.5016
Most common pitch class prevalence - 0.2929

Average note duration – 0.9721
Dominant spread - 3.4231
Tempo - 80.3077
Key mode - 1.5385
Climax position - 0.4247
Consecutive identical pitches - 3.9058
Note density - 11.1158
Spectral texture MFCC 4 - -2.1943
Average time between attacks - 0.3116
Variability of note duration - 1.3209
Variability of time between attacks - 0.2322
Note prevalence fretless bass - 0.0083
String ensemble fraction - 0.1818
Saxophone fraction – 0
Spectral texture MFCC 2 - 9.9105
Average duration accent - 0.8123
Polyrhythms - 0.5567
Staccato incidence - 0.1441
Repeated notes - 0.1016
Average melodic complexity - 5.0149
Most common pitch class prevalence - 0.2546

Average note duration – 0.5089
Dominant spread - 4.2308
Tempo - 106.3846
Key mode - 1.3846
Climax position - 0.4674
Consecutive identical pitches - 6.7862
Note density - 15.3345
Spectral texture MFCC 4 - -1.1140
Average time between attacks - 0.1925
Variability of note duration - 0.5640
Variability of time between attacks - 0.1527
Note prevalence fretless bass - 0.0218
String ensemble fraction - 0.0531
Saxophone fraction - 0.0183
Spectral texture MFCC 2 - 14.8053
Average duration accent - 0.8013
Polyrhythms - 0.5211
Staccato incidence - 0.2233
Repeated notes - 0.1518
Average melodic complexity - 5.1614
Most common pitch class prevalence - 0.2511

Fig. 6. Mean values of relevant features of musical samples for each affective quadrant

VII. Conclusion

We presented an extension of a previous work that undertook music emotion classification as a regression problem. SVM regression obtained the best results in the prediction and classification of the dimensions of valence and arousal. Validation results using the coefficient of determination confirmed that the prediction/classification of arousal (90.31%/81.55%) is easier than the prediction/classification of valence (89.37%/79.86%). Rhythmic (e.g., tempo, note density and average/variation of note duration), melodic (e.g., climax position and melodic complexity) and textural (e.g., spectral texture MFCCs) features proved to be very important to valence and arousal. Harmonic (e.g., key mode) and dynamics features (e.g., staccato incidence) were also important to predict, respectively, the valence and arousal. A correlation coefficient of 62.95% was obtained between valence and arousal.

With these satisfactory results, we feel ready to move to the second stage of our work, that consists in transformation of the affective content of selected music to approximate even further its affective content to an intended emotion. Both the selection and transformation will use the obtained information stored in the knowledge base (Fig. 4, 5, 6).

References

[1] Cambouropoulos, E. "The local boundary detection model (lbdm) and its application in the study of expressive timing.", *Int. Computer Music Conf.*, 2001.

[2] Eerola, T., Toiviainen, P. "Mir in matlab: The midi toolbox.", *International Conference on Music Information Retrieval*, 2004.

[3] Gabrielsson, A., Lindström, E. "The Influence Of Musical Structure On Emotional Expression." *Music and emotion: Theory and research*. Oxford University Press, 2001, 223–248.

[4] Korhonen, M. "Modeling continuous emotional appraisals of music using system identification." *Master's thesis*, University of Waterloo, 2004.

[5] Kuo, F., Chiang, M., Shan, M., Lee, S. "Emotion-based music recommendation by association discovery from film music." *ACM International Conference On Multimedia*, 2005, 507–510

[6] Li, T., Ogihara, M. "Detecting emotion in music." *International Conference on Music Information Retrieval*, 2003, 239–240

[7] Lindstrom, e. "A Dynamic View of Melodic Organization and Performance." *PhD thesis*, Acta Universitatis Upsaliensis Uppsala, 2004.

[8] Livingstone, S., Brown, A. "Dynamic response: real-time adaptation for music emotion." *Australasian Conference On Interactive Entertainment*, 2005, 105–111.

[9] McKay, C., Fujinaga, I. "jsymbolic: A feature extractor for midi files." *International Computer Music Conference*, 2006.

[10] Muyuan, W., Naiyao, Z., Hancheng, Z. "User-adaptive music emotion recognition." *International Conference on Signal Processing*, 2004.

[11] Oliveira, A., Cardoso, A. "Towards affective-psychophysiological foundations for music production." *Affective Computing and Intelligent Interaction*, 2007, 511–522.

[12] Oliveira, A., Cardoso, A. "Towards bi-dimensional classification of symbolic music by affective content." *International Computer Music Conference (to be published)*, 2008.

[13] Russell, J.A. "Measures of emotion." *Emotion: Theory, research, and experience*, 1989, 4, 83-111.

[14] Schubert, E. "Measurement and Time Series Analysis of Emotion in Music." *PhD thesis*, University of New South Wales, 1999.

[15] Sorensen, A. and Brown, A. "Introducing jMusic." *Australasian Computer Music Conference*, 2000, 68-76.

[16] Thimm K. and Fischer B. "Emotional responses to music: Influence of psycho-acoustical features." *National Conferences on Undergraduate Research*, 2003.

[17] van de Laar, B. "Emotion detection in music, a survey." *Twente Student Conference on IT*, 2006.

[18] Witten, I., Frank, E., Trigg, L., Hall, M., Holmes, G., Cunningham, S. "Weka: Practical machine learning tools and techniques with java implementations." *International Conference on Neural Information Processing*, 1999, 192–196.

[19] Wu, T., Jeng, S.: "Extraction of segments of significant emotional expressions in music." *Workshop on Computer Music and Audio Technology*, 2006.

Mode-dependent Differences in Chord Classification under an Original Computational Method of Tonal Structure Analysis

Miroslaw Majchrzak, PhD student
Institute of Art, Polish Academy of Sciences in Warsaw, Poland
mmajchrzak77@wp.pl

Abstract — **Basing upon original computational analytic method (Majchrzak 2005, 2007), the present work aims, at: 1) Showing the differences for the major key and the minor (harmonic) key in the classification of chords, as an aspect of importance for interpreting a piece's tonal structure diagram; 2) Drawing attention to the subordination of the minor key as versus the major key in the chord classification, using the same algorithm. The relations between chords appearing in the major and minor (harmonic) key are shown by applying the comparisons of: 1) third-based chords; 2) degrees in the C major and A minor keys, on which the same diatonic chords appear.**

Keywords: tonality, major key, minor key, analysis

I. Introduction

The invention of harmony in the baroque period was one source of polemics around music in 17th and 18th centuries. Among those who investigated the foundations of harmony on a philosophical basis, including the legitimacy of the two modes, are the founders of new scientific methods (Kepler 1619, Mersenne 1637, or, Descartes 1650). In the same period, both modes (i.e. major and minor) tend to be reduced to a single, i.e. major, scale – in that a minor is but a variety of the 'perfect' major scale. Theoretical works on scale modes, justifying the existence of scales, show minor scales as subordinate to the major. In Helmholtz's approach, the minor scale is not part of the music's beauty; nor can it be classed under the natural or rational system. Also Rameau (1722) was of opinion only the existence of the major mode is explainable in rational terms in the world of harmony. He considered the minor mode an unnatural variety of the major mode.

Our contemporary theoretical works on harmony, tonality, methods of main key determination in a musical piece, maps of chord relations, etc., are indicative of certain problems with the minor key (Shepard 1982, Chew 2000, Krumhansl 1990, Honingh 2007). The minor key issue also concerns the Author's computational method of analysis of the tonal structure in pieces of music (Majchrzak 2005, 2007).

II. Method of Analysis of the Tonal Structure

Original method consists in assignation of chords appearing in a piece of music to individual key ranges being keys in their respective natural variety. Using the analytical method in question, a diagram of tonal structure of a piece can be produced, such tonal structure being understood as quantitative relation of key ranges for which specific chords have been classified. We mark the keys with the consecutive integers: the sharp keys with positive numbers, the flat keys – with negative numbers. The absolute value of the integer designates the number of accidentals in the key. The number (3) marks the keys of A major and F sharp minor (natural); the number (-1) – the keys of F major and D minor (natural).

For any tone, we can determine the keys it appears in. For instance, the tone D appears in these keys: (-3, -2, -1, 0, 1, 2, 3)[1]. The tone E appears in the following keys: (-1, 0, 1, 2, 3, 4, 5)[2]. The tone C appears in these keys: (-5, -4, -3, -2, -1, 0, 1)[3]. This is similarly so for any and each chord. For example, the tones of the C major chord appears in the following keys: (-5, -4, -3, -2, -1, 0, 1), (-1, 0, 1, 2, 3, 4, 5), (-4, -3, -2, -1, 0, 1, 2).

Axioms:

1) In the event that one of the chord tones is an octave transposition of another, then, such a tone shall not be taken into account whilst classifying the chord;

2) Any tone being distant from one another by one or more octaves shall be approached on an equivalent basis.

The substratum for our chord classification is the arithmetic average of keys wherein the tones of a given diatonic chord appear:

$$\text{arithmetic average} = (x_1 + x_2 + x_3 + \ldots + x_n) / n$$

$x_1 + x_2 + x_3 + \ldots + x_n$ – keys wherein the tones of a given diatonic chord appear; n – number of all keys.

Examples:

1) DFsharp:

$$\text{AA (arithmetic average)} = 2$$

$$\frac{(-3-2-1+0+1+2+3)+(1+2+3+4+5+6+7)}{7+7}$$

[1] (E flat major and C minor, B flat major and G minor, F major and D minor, C major and A minor, G major and E minor, D major and B minor, A major and F sharp minor).

[2] (F major and D minor, C major and A minor, G major and E minor, D major and B minor, A major and F sharp minor, E major and C sharp minor, B major and G sharp minor).

[3] (D flat major and B flat minor, A flat major and F minor, E flat major and C minor, B flat major and G minor, F major and D minor, C major and A minor, G major and E minor).

2) BDFA:

$$AA = 0{,}25$$

$$\frac{(0+1+2+3+4+5+6)+(-3+2+1+0+1+2+3)+(-6-5-4-3-2-1+0)+(-2-1+0+1+2+3+4)}{7+7+7+7}$$

3) DFsharpAC:

$$AA = 0{,}75$$

$$\frac{(-3-2-1+0+1+2+3)+(1+2+3+4+5+6+7)+(-2-1+0+1+2+3+4)+(-5-4-3-2-1+0+1)}{7+7+7+7}$$

4) GCsharp:

$$AA = 2$$

$$\frac{(-4-3-2-1+0+1+2)+(2+3+4+5+6+7+8)}{7+7}$$

In this method:

Arithmetic average space:

– all numeric values derivable from the above arithmetic-average formula. The arithmetic average space is divided into key ranges (KRs), each of which is a key range with a given number of clef signs. E.g., the key range of one-flat keys (F major and D minor) encompasses the arithmetic average space's open-ended range, spanning between -0,5 and -1,5. The key range of two-sharp keys (D major and B minor) encompasses the arithmetic average spanning between 1,5 and 2,5. The key range of four-sharp keys (E major and C# minor) encompasses the arithmetic average spanning between 4,5 and 5,5.

Chords and Key Range

Examples: the chord GBDF (AA = -0,25) belongs to KR 0. The chord EG#BC# (AA = 4) belongs to KR 4.

2KRs chord

– any chord whose arithmetic average belongs to two adjacent KRs. E.g., the arithmetic average of the CEGB chord is 0,5; the chord belongs to both KR 0 (C major and A minor) and KR 1 (G major and E minor). The arithmetic average of the CDEFGA chord is -0,5; the chord belongs to both KR -1 (F major and D minor) and KR 0 (C major and A minor).

N-D

– non-diatonic chords.

Analysis of pieces can be displayed in the form of diagram.

Example**:**

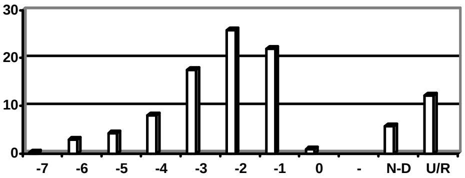


Fig. 1. Chopin, Mazurka B flat major, Op. 17, No. 1

Where:

Horizontal Axis:

Key ranges. For example: -4 (Key range of A flat major and F minor), -2 (Key range of B flat major and D minor), 0 (Key ranges of C major and A minor)

Vertical Axis:

Percentage domination of given key ranges

III. Third-based chords built on individual degrees of C major/A minor keys

As discussed hereinabove, the analytical method consists in assignment of diatonic chords to individual ranges of a key, which is followed by a quantitative comparison of the key ranges. Let us take a look at the differences in assignment to key ranges of triad appearing on individual grades of C major and A minor keys.

1) C major:

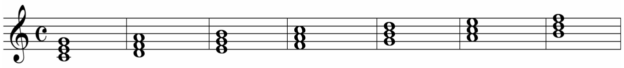


Fig. 2. Triads appearing on individual grades of C major key.

KR 0 KR -1 KR 1 KR -1 KR 1 KR 0 KR 0

2) A minor:

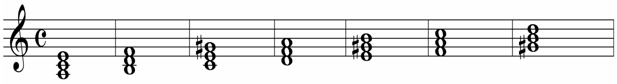


Fig. 2. Triads appearing on individual grades of A minor (harmonic).

KR 0 KR 0 N-D KR -1 KR 4 KR -1 KR 3

We could see above that all the triads built upon individual C major key grades are assigned to the key's three ranges, including: KR 0 (C major and A minor), KR -1 (F major and D minor), KR 1 (G major and E minor). Triads created on individual grades of the A minor harmonic key belong to the key's four ranges whilst one of them belongs to the N-D group. Now, let us have a closer look at third-based chords built up on individual C major and A minor (harmonic) key grades.

TABLE I.
Third-based chords built on individual degrees of C major and A minor (harmonic) keys.

Dyads				
	C major		A minor	
Degree	**Chord**	**Key range**	**Chord**	**Key range**
1st	CE	KR 0	AC	KR -1, KR 0
2nd	DF	KR -2, KR -1	BD	KR 1, KR 2
3th	EG	KR 0, KR 1	CE	KR 0
4th	FA	KR -1	DF	KR -2, KR -1
5th	GB	KR 1	EG#	KR 4
6th	AC	KR -1, KR0	FA	KR -1
7th	BD	KR 1, KR 2	G#B	KR 4, KR 5

Triads				
	C major		A minor	
Degree	**Chord**	**Key range**	**Chord**	**Key range**
1st	CEG	KR 0	ACE	KR 0
2nd	DFA	KR -1	BDF	KR 0
3th	EGB	KR 1	CEG#	N-D

4th	FAC	KR -1	DFA	KR -1
5th	GBD	KR 1	EG#B	KR 4
6th	ACE	KR 0	FAC	KR -1
7th	BDF	KR 0	G#BD	KR3

Four-note chords				
	C major		**A minor**	
Degree	**Chord**	**Key range**	**Key range**	**Key range**
1st	CEGB	KR 0, KR 1	ACEG#	N-D
2nd	DFAC	KR -1	BDFA	KR 0
3th	EGBD	KR 1	CEG#B	N-D
4th	FACE	KR -1, KR 0	DFAC	KR -1
5th	GBDF	KR 0	EG#BD	KR3
6th	ACEG	KR 0	FACE	KR -1, KR 0
7th	BDFA	KR 0	G#BDF	N-D

Five-note chords				
	C major		**A minor**	
Degree	**Chord**	**Key range**	**Key range**	**Key range**
1st	CEGBD	KR 0	ACEG#B	N-D
2nd	DFACE	KR 0	BDFAC	KR 0
3th	EGBDF	KR 0	CEG#BD	N-D
4th	FACEG	KR -1	DFACE	KR 0
5th	GBDFA	KR 0	EG#BDF	N-D
6th	ACEGB	KR 1	FACEG#	N-D
7th	BDFAC	KR 0	G#BDFA	N-D

Six-note chords				
	C major		**A minor**	
Degree	**Chord**	**Key range**	**Key range**	**Key range**
1st	CEGBDF	KR 0	ACEG#BD	N-D
2nd	DFACEG	KR -1, KR 0	BDFACEG#	N-D
3th	EGBDFA	KR 0	CEG#BD	N-D
4th	FACEGB	KR 0	DFACEG#	N-D
5th	GBDFAC	KR 0	EG#BDFA	N-D
6th	ACEGBD	KR 0, KR 1	FACEG#B	N-D
7th	BDFACE	KR 0	G#BDFAC	N-D

A. TRIADS

1st, 4th:

Chords based on 1st degree of C major and A minor keys are classed in the key range where they function as the keynotes, i.e. KR 0 (C major, A minor). The situation where triads built upon the same degree in the keys C major and A minor are part of a single KR is to be met only once: this concerns chords built on the fourth degree. The F major and D minor chords are part of PT -1. The situation is different for triads built on the remaining degrees of those keys.

2nd:

The chord built up on the 2nd degree of the C major – i.e. the subdominant of the 2nd degree belongs, as shown above, to KR -1, the range to which the C major subdominant chord belongs as well. In the A minor, the chord built on the 2nd degree is part of the same key range as the chord built on 1st degree (i.e. the minor keynote), that is, KR 0. Having said that, why should the triad built on the 2nd degree of A minor key belong to PT 0? The BDF chord may be considered as a dominant seventh without the root in C major key. It then appears in the key range within which the chord appears into which it is resolved (according to the classic theory of harmony, the BDF chord may be resolved to the C major chord).

3th:

The chord built up on the 3rd degree of the C major, composed of EGB tones, belongs to KR 1, and so, to the KR where the G major chord appears. The chord on the 3rd degree in the A minor is an augmented chord, which means that it is not assigned to a key range. Instead, it is classed under a separate group of non-diatonic chords (N-D).

5th:

The triad built on the 5th degree (the dominant) in the C major, i.e. the G major chord, is classified in KR 1. The range is situated right of the keynote's range (KR 0 in the C major). In A minor key, the minor keynote appears in KR 0. In turn, the chord built on the 5th degree of the A minor, i.e. the major dominant, is classified as KR 4. Then, how should the dominant's situation be explained, in a range fixed as many as four ranges away from the range where the minor keynote in the A minor is classed? The E major chord may act as a keynote for the E major key. Hence, it is contained within KR 4, similarly as the C major chord in KR 0 or the A flat major chord KR -4.

6th:

In the C major, the sixth-grade keynote (ACE) appears within the same key range as the keynote (CEG), i.e. KR 0. In the A minor, the triad built on the sixth degree is situated in KR -1, that is, a range located left of the range wherein the keynote chord appears.

7th:

In both C major and A major key, a diminished chord appears upon degree 7th. We have already come across the BDF chord on the grounds of A minor key (as its 2nd-degree chord). As for the GsharpBD chord appearing on the 7th degree of the A minor, the following question arises: How can we interpret the position of a chord built up on the 7th degree of A minor key in the range of A major and F sharp minor keys (KR 3)? The chord composed of the notes GsharpBD may be deemed to be the dominant seventh without the root for A major key, i.e. KR 3. Thus, the second of the diminished triads built upon the A minor degrees better corresponds with the major key (A major – KR 3) than with a minor one (A minor – KR 0).

B. FOUR-NOTE CHORDS

1st:

The arithmetic average of the keys where the notes appear of the four-part chord built on 1st degree of the C major (i.e. CEGB) equals 0,5. Thus, the chord belongs to both KR 0 and KR 1[4]. As for the minor key, one has to do with

[4] 2KRs.

a chord whose notes are not reducible to a single key of the natural variety (BDFGsharp)[5], and hence, we will not take it into consideration for the purpose of assignment to individual key ranges.

2nd:

The four-note chord built on the 2nd degree of the C major belongs to KR -1. Thich chord can be considered as the minor keynote with a small seventh added in the D minor, or, the keynote with a great sixth added in the F major. The four-part chord BDFA appearing on degree 2nd of the A minor is classed under KR 0. Within this same range, C major chord appears, to which BDFA chord, being the C major key's dominant ninth without the root, gets most frequently resolved.

3rd

The structure of four-note chord EGBD is identical to that of DFAC chord, whereas the CEGsharpB chord is put in our breakdown in a separate column (N-D), and not assigned to an individual key range.

4th:

In the C major, the four-note chord built on degree 4th has a structure identical as the four-note chord built on the key's 1st degree (the chord belongs to both KR -1 and KR 0). In the A minor, the four-note chord built on the 4th degree is of an identical structure as the one built on degree 1st of the C major (the chord belongs to KR -1).

5th:

The GBDF chord is part of KR 0, i.e. to that major-key range in which it operates as the dominant seventh. In the case of the A minor dominant seventh, we come across a troublesome case. This chord does not, namely, belong to the key range wherein the A minor keynote appears (i.e. KR 0), but is part of KR 3 instead. A similar situation was the case when it came to discussing the GsharpBD chord. The GBDF chord belongs to the range where the major tonic appears (CEG chord appearing in KR 0), whereas it is not part of the range where the major tonic appears to which this chord can be resolved (CEflatG appears in KR -3). A similar thing happens with EGsharpBD chord, which appears in the range where the major tonic is classed to which it is resolved, that is, in KR 3. (The A major chord to which EG#BD chord gets resolved appears in KR 3).

6th, 7th:

Chords build like: ACEG, FACE, BDFA have already been discussed. The diminished four-part chord is a non-diatonic chord (N-D).

C. Five-note chords

Most five-note chords built on A minor key degrees are part of the N-D group. In the C major, most of the chords belong to the key's main range, and two of them belong to KR -1 and KR 1, respectively.

D. Six-note chords

All the six-note chords created on individual degrees of the A minor, in its harmonic variety, contain an augmented four-part chord. This means that these are not assigned to key ranges (N-D). As for the C major, all the six-part chords belong to KR 0. Two of them, created upon degrees 2nd and 6th, respectively, are 2KRs chords.

E. Dyads

To end with, let us take a look on third-sized dyads. As it can be seen in the table above, the differences are remarkable also for the dyads.

IV. Diatonic chords of varied structure

The previous chapter discussed third-based chords built upon individual C major and A minor keys' degrees. Some of the minor-key chords were unclassified with respect to the key ranges, as their tones could not be reduced to a single natural key. Now, let us turn attention to diatonic chords with a diverse interval structure. The subsequent table lines specify chords belonging to KR 0 (C major, A minor) and the C major/A minor degrees whereupon the chords are created. Example: Let us take any triad, e.g. CDE. The arithmetic average equals 0, so the triad is contained within KR 0 (C major and A minor).

C major key degrees:				1	2	3	4	5	6	7	
Tones:	G	A	B	**C**	**D**	**E**	F	G	A	B	C
A minor key degrees:	1	2	3	4	5	6	7				

In the C major, it is built on degrees 1st, 2nd and 3rd. These degrees appear more important than those in the case of A minor key (3rd, 6th 5th). The first three notes of the C major comprise the tonic's prime and third, whereas in the A minor, these are the keynote's third and fifth with an added fourth.

As for 2KRs chords, two identically structured chords will be quoted in our tables: the first belonging to KR 0 and KR 1:

background:
░

and the second, to KR -1 and KR 0:
background:
▒

[5] N-D.

TABLE II.
DIATONIC CHORDS WITH A DIVERSE INTERVAL STRUCTURE.

Dyads		
Chord	**C major key degrees**	**A minor key degrees**
G A B C D E F G A B C	**1, 3** (CE)	**3, 5** (CE)
G A B C D E F G A B C	**3, 5** (EG)	**5, 7** (EG)
G A B C D E F G A B C	**1, 6** (CA)	**1, 3** (AC)
G A B C D E F G A B C	**5, 6** (GA)	**1, 7** (AG)
G A B C D E F G A B C	**1, 7** (CB)	**2, 3** (BC)
G A B C D E F G A B C	**3, 4** (EF)	**5, 6** (EF)
G A B C D E F G A B C	**4, 7** (FB)	**2, 6** (BF)
G A B C D E F G A B C	**2, 6** (DA)	**1, 5** (AD)
G A B C D E F G A B C	**2, 5** (DG)	**4, 7** (DG)

Triads		
Chord	**C major key degrees**	**A minor key degrees**
G A B C D E F G A B C	**1, 3, 5** (CEG)	**3, 5, 7** (CEG)
G A B C D E F G A B C	**1, 3, 6** (CEA)	**1, 3, 5** (ACE)
G A B C D E F G A B C	**2, 5, 6** (DGA)	**1, 4, 7** (ADG)
G A B C D E F G A B C	**2, 4, 7** (DFB)	**2, 4, 6** (BDF)
G A B C D E F G A B C	**4, 6, 7** (FAB)	**1, 2, 6** (ABF)
G A B C D E F G A B C	**4, 5, 7** (FGB)	**2, 6, 7** (BFG)
G A Bb C D E F G A Bb C	**3, 6, 7b** (EABb)	**1, 2b, 5** (ABbE)
G A B C D E F# G A B C	**1, 4#, 5** (CF#G)	**3, 6#, 7** (CF#G)
G A B C D E F G A B C	**1, 2, 6** (CDA)	**1, 3, 4** (ACD)
G A B C D E F G A B C	**2, 3, 5** (DEG)	**4, 5, 7** (DEG)
G A B C D E F G A B C	**1, 5, 7** (CGB)	**2, 3, 7** (BCG)
G A B C D E F G A B C	**3, 4, 6** (EFA)	**1, 5, 6** (AEF)
G A B C D E F G A B C	**1, 2, 3** (CDE)	**3, 4, 5** (CDE)
G A B C D E F G A B C	**2, 3, 4** (DEF)	**4, 5, 6** (DEF)
G A B C D E F G A B C	**1, 2, 7** (CDB)	**2, 3, 4** (BCD)

Four-note chords		
Chord	**C major key degrees**	**A minor key degrees**
G A B C D E F G A B C	**2, 4, 5, 7** (DFGB)	**2, 4, 6, 7** (BDFG)
G A B C D E F G A B C	**2, 4, 6, 7** (DFAB)	**1, 2, 4, 6** (ABDF)
G A B C D E F G A B C	**2, 3, 5, 6** (DEGA)	**1, 4, 5, 7** (ADEG)
G A B C D E F G A B C	**1, 2, 5, 6** (CDGA)	**1, 3, 4, 7** (ACDG)
G A B C D E F G A B C	**1,3, 5, 6** (CEGA)	**1, 3, 5, 7** (ACEG)
G A B C D E F G A B C	**1, 3, 5, 7** (CEGB)	**2, 3, 5, 7** (BCEG)
G A B C D E F G A B C	**1, 3, 4, 6** (CEFA)	**1, 3, 5, 6** (ACEF)
G A B C D E F G A B C	**1, 2, 3, 6** (CDEA)	**1, 3, 4, 5** (ACDE)
G A B C D E F G A B C	**1, 2, 3, 5** (CDEG)	**3, 4, 5, 7** (CDEG)
G A B C D E F# G A B C	**1, 2, 4#, 5** (CDF#G)	**3, 4, 6#, 7** (CDF#G)
G A Bb C D E F G A Bb C	**2, 3, 6, 7b** (DEABb)	**1, 2b, 4, 5** (ABbDE)
G A B C D E F G A B C	**1, 4, 6, 7** (CFAB)	**1, 2, 3, 6** (ABCF)
G A B C D E F G A B C	**3, 4, 5, 7** (EFGB)	**2, 5, 6, 7** (BEFG)
G A B C D E F G A B C	**1, 2, 5, 7** (CDGB)	**2, 3, 4, 7** (BCDG)
G A B C D E F G A B C	**4, 5, 6, 7** (FGAB)	**1, 2, 6, 7** (ABFG)
G A B C D E F G A B C	**1, 3, 4, 7** (CEFB)	**2, 3, 5, 6** (BCEF)
G A B C D E F G A B C	**2, 3, 4, 7** (DEFB)	**2, 4, 5, 6** (BDEF)
G A Bb C D E F G A Bb C	**3, 5, 6, 7b** (EGABb)	**1, 2b, 5, 7** (ABbEG)
G A B C D E F# G A B C	**1, 4#, 5, 6** (CF#GA)	**1, 3, 6#, 7** (ACF#G)
G A B C D E F G A B C	**1, 2, 4, 7** (CDFB)	**2, 3, 4, 6** (BCDF)
G A B C D E F G A B C	**1, 5, 6, 7** (CGAB)	**1, 2, 3, 7** (ABCG)

G	A	B	C	D	E	F	G	A	B	C	**1, 2, 6, 7** (CDAB)	**1, 2, 3, 4** (ABCD)
G	A	B	C	D	E	F	G	A	B	C	**2, 3, 4, 5** (DEFG)	**4, 5, 6, 7** (DEFG)

Five-note chords												
Chord											**C major key degrees**	**A minor key degrees**
G	A	B	C	D	E	F	G	A	B	C	**1, 2, 3, 5, 6** (CDEGA)	**1, 3, 4, 5, 7** (ACDEG)
G	A	B	C	D	E	F	G	A	B	C	**2, 4, 5, 6, 7** (DFGAB)	**1, 2, 4, 6, 7** (ABDFG)
G	A	Bb	C	D	E	F	G	A	Bb	C	**2, 3, 5, 6, 7b** (DEGABb)	**1, 2b, 4, 5, 7** (ABbDEG)
G	A	B	C	D	E	F#	G	A	B	C	**1, 2, 4#, 5, 6** (CDF#GA)	**1, 3, 4, 6#, 7** (ACDF#G)
G	A	B	C	D	E	F	G	A	B	C	**1, 2, 3, 4, 6** (CDEFA)	**1, 3, 4, 5, 6** (ACDEF)
G	A	B	C	D	E	F	G	A	B	C	**1, 2, 3, 5, 7** (CDEGB)	**2, 3, 4, 5, 7** (BCDEG)
G	A	B	C	D	E	F	G	A	B	C	**1, 3, 4, 6, 7** (CEFAB)	**1, 2, 3, 5, 6** (ABCEF)
G	A	B	C	D	E	F	G	A	B	C	**1, 3, 4, 5, 7** (CEFGB)	**2, 3, 5, 6, 7** (BCEFG)
G	A	B	C	D	E	F	G	A	B	C	**2, 3, 4, 5, 7** (DEFGB)	**2, 4, 5, 6, 7** (BDEFG)
G	A	B	C	D	E	F	G	A	B	C	**1, 2, 4, 6, 7** (CDFAB)	**1, 2, 3, 4, 6** (ABCDF)
G	A	B	C	D	E	F	G	A	B	C	**1, 4, 5, 6, 7** (CFGAB)	**1, 2, 3, 6, 7** (ABCFG)
G	A	B	C	D	E	F	G	A	B	C	**3, 4, 5, 6, 7** (EFGAB)	**1, 2, 5, 6, 7** (ABEFG)
G	A	B	C	D	E	F	G	A	B	C	**1, 2, 3, 4, 7** (CDEFB)	**2, 3, 4, 5, 6** (BCDEF)
G	A	B	C	D	E	F	G	A	B	C	**2, 3, 4, 5, 6** (DEFGA)	**1, 4, 5, 6, 7** (ADEFG)
G	A	B	C	D	E	F	G	A	B	C	**1, 3, 4, 5, 7** (CEFGB)	**2, 3, 5, 6, 7** (BCEFG)

Six-note chords												
Chord											**C major key degrees**	**A minor key degrees**
G	A	B	C	D	E	F	G	A	B	C	**1, 2, 4, 5, 6, 7** (CDFGAB)	**1, 2, 3, 4, 6, 7** (ABCDFG)
G	A	B	C	D	E	F	G	A	B	C	**2, 3, 4, 5, 6, 7** (DEFGAB)	**1, 2, 4, 5, 6, 7** (ABDEFG)
G	A	B	C	D	E	F	G	A	B	C	**1, 2, 3, 5, 6, 7** (CDEGAB)	**1, 2, 3, 4, 5, 7** (ABCDEG)
G	A	B	C	D	E	F	G	A	B	C	**1, 2, 3, 4, 5, 6** (CDEFGA)	**1, 3, 4, 5, 6, 7** (ACDEFG)
G	A	B	C	D	E	F	G	A	B	C	**1, 3, 4, 5, 6, 7** (CEFGAB)	**1, 2, 3, 5, 6, 7** (ABCEFG)
G	A	B	C	D	E	F	G	A	B	C	**1, 2, 3, 4, 6, 7** (CDEFAB)	**1, 2, 3, 4, 5, 6** (ABCDEF)
G	A	B	C	D	E	F	G	A	B	C	**1, 2, 3, 4, 5, 7** (CDEFGB)	**2, 3, 4, 5, 6, 7** (BCDEFG)

V.CONCLUSIONS

The examples discussed above enable us to draw attention to the differences in the assignment of chords to the key ranges, conditional upon the key's mode.

Basing upon the examples quoted (C major and A minor keys, in our case), tentative conclusions may be drawn with respect to a superiority of the major key over the minor: 1) In major key, dominant forms are frequently contained within the key range in which the major tonic appears to which they are resolvable. In the minor, dominant forms are distant from the range wherein the minor chord (minor keynote) appears to which they are resolved. 2) In minor key (harmonic variety), chords appear that are not assignable to key ranges (which also refers to dominant forms, e.g. dominant ninth with a small ninth, or, diminished four-part chord).

ACKNOWLEDGMENT

I am grateful to Dr Alan Marsden, Lancaster University, for helpful comments. I thank also Tristan Korecki for translation assistance.

REFERENCES

[1] E. Chew (2000), *Towards a Mathematical Model of Tonality*, Massachusetts Institute of Technology, Dissertation.

[2] R. Descartes (1650), *Musicae compendium*, Utrecht 1650.

[3] A. Honingh (2007), *Automatic modulation finding using convex sets of notes*, 1st International Conference of Society for Mathematics and Computation in Music, Berlin, May 18-20, 2007.

[4] J. Kepler (1619), *Harmonices mundi* 1619.

[5] C. L. Krumhansl (1990), *Cognitive foundations of musical pitch*. Oxford University Press.

[6] M. Mersenne (1637-1637), *Harmonie universelle*, Paris 1636-1637.

[7] M. Majchrzak (2005), *Divergences and convergences of major and minor key distribution series in musical pieces of the tonal harmony supremacy period*, Master Thesis, Wroclaw Academy of Music 2005, (supervisor: prof. dr Stanislaw Krupowicz).

[8] M. Majchrzak (2007), *Irrelative System in Tonal Harmony*, 1st International Conference of Society for Mathematics and Computation in Music, Berlin, May 18-20, 2007.

[9] M. Majchrzak (2007), *The Tonal Structure of Chopin' s Mazurkas*, Presentation at International Musicological Society, Study Group on Musical Data and Computer Applications. Zurich 2007.

[10] J. P. Rameau (1722), *Traité de l'harmonie réduite à son principe naturel*, Paris 1722.

[11] R. N. Shepard (1982), Geometrical approximations to the structure of musical pitch, Psychological Review 89.

Concatenative Synthesis of Expressive Saxophone Performance

Stefan Kersten*, Rafael Ramirez*
*Music Technology Group, Universitat Pompeu-Fabra, Barcelona, Spain

Abstract—**In this paper we present a systematic approach to applying expressive performance models to non-expressive score transcriptions and synthesizing the results by means of concatenative synthesis. Expressive performance models are built from score transcriptions and recorded performances by means of decision tree rule induction, and those models are used both to transform inexpressive input scores and to guide the concatenative synthesizer unit selection.**

I. Introduction

In the past, important approaches to expressive performance modeling have been empirical methods based on statistical analysis, mathematical modeling, and "analysis-by-synthesis" (see the summary provided in [1]). In all these approaches, it is a person who is responsible for devising a theory or a mathematical model which captures different aspects of musical expressive performance. The theory or model is later tested on real performance data in order to determine its accuracy.

Our approach as well as the one described in [1] is based on building computational models of expressive performance by machine learning, in our case inductive logic decision tree models. Those models are used to predict expressive transformations for inexpressive scores as well as guiding sample database note selection and transition modeling in a concatenative saxophone synthesizer.

II. Expressive performance analysis

Figure 1 gives an overview of the functional parts of the expressive performance analysis and synthesis system.

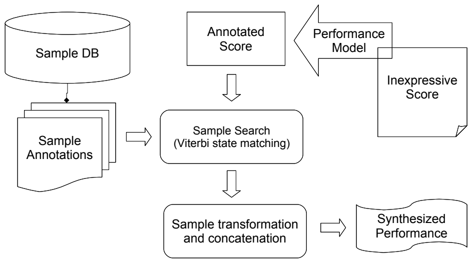


Fig. 1. Expressive performance analysis and synthesis system overview.

Modeling of expressive performance involves analysis of recorded performances of a musical piece and the comparison of the symbolical information extracted with the one present in the score in order to build a computational model for a particular performer in a particular style of music.

Our approach can be divided into three distinct steps: In a preprocessing stage low- and high-level perceptive features are extracted from a short-time fourier transform (STFT) representation of the musical audio recording and are grouped into a note-level transcription (Fig. 2). The transcribed representation is aligned to the symbolic score in a second step and finally a computational model of the performance's characteristics is built.

A. Note segmentation and feature extraction

In this section we shall be concerned with extracting the symbolic note-level descriptors from a performance recording that are needed both for building the computational performance model and for synthesizing a score enriched with expressivity annotations.

Most of the features are calculated from a short-time fourier transition (STFT) signal representation, i.e. overlapping frames of time-domain audio data that are multiplied by a window function and transformed to the frequency domain by the discrete fourier transform (DFT). In our descriptor database we used a frame size of 1024 with an overlap of 50% at a sample rate of 44100 kHz. The window function used is a Kaiser-Bessel window [2] with a 25dB side-lobe to main-lobe ratio.

The main low-level features used for describing expressive performance are fundamental frequency and mean energy. Log-mean-energy is extracted from a time domain representation using frames of size 1024 with a 50% overlap weighted by a Blackman-Harris window w of length N according to (1).

$$E_n = 20\log\frac{\sum_{m=-\infty}^{m=\infty}[x(m)w(n-m)]^2}{N} \tag{1}$$

Instantaneous fundamental frequency is extracted following the two-way mismatch procedure described in [3] and [4]. In order to obtain a "brightness" measure, the spectral centroid of the discrete spectrum $X(n)$ of size N is extracted according to (2).

$$F_{centroid} = \frac{\sum_{n=0}^{n=N-1} f(n)|X(n)|}{\sum_{n=0}^{n=N-1}|X(n)|} \tag{2}$$

Note segmentation is performed in a two-step algorithm based on descriptors extracted from the spectral signal representation. First, an onset function is calculated based on energy envelopes in different bands and by applying

Note onset
Note offset
Note duration
Fundamental Frequency
Mean energy
Energy envelope attack time
Energy envelope sustain init level
Energy envelope sustain end level
Energy envelope sustain time
Energy envelope release time
Legato left
Legato right
Mean spectral centroid

TABLE I
FEATURES EXTRACTED DURING PRE-PROCESSING

psycho-acoustical knowledge [5]. In a second step, the onset function is combined with a pitch-transition function to yield the final note onset and offset detection function.

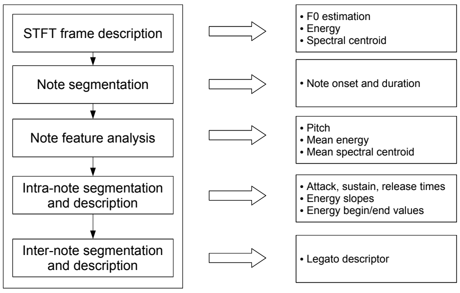


Fig. 2. Feature extraction for Expressive Performance Analysis.

The note-level feature j is calculated from the frame-based feature values by calculating the mean value over the analysis frame indices contained in the ith note's onset and offset times expressed in terms of analysis frames k_{on} and k_{off} (3). In the case of note fundamental frequency and pitch estimation a histogramming approach according to [6] is employed, in order to smooth out large errors in instantaneous fundamental frequency extraction.

$$\overline{F_j}(i) = \frac{\sum_{k=k_{\text{on}}}^{k_{\text{off}}} F_j(k)}{k_{\text{off}} - k_{\text{on}}} \tag{3}$$

Once note onsets and offsets have been determined, energy envelope attack and release times and the corresponding energy levels are extracted along with a *legato* descriptor, that captures the "smoothness" of transition between two successive notes [7]. Table I lists all of the features being used either by the model generation step, the concatenative synthesizer or both.

Notes –or *units*– extracted from performance audio files are organized in a file-system based database containing phrase- and note-level information in XML files and accompanying PCM audio files and binary analysis files used by the synthesizer.

B. Modeling of expressive performance

The symbolic representation obtained in the analysis step is aligned to the symbolic score by dynamic timewarping and subsequent manual resolution of errors. Since a score note only contains a limited amount of useful information for expressivity analysis, an attempt was made to capture and associate more meaningful musical context with each individual score note. For this purpose the Implication/Realization model described in [8] is used, which describes relationships within a set of notes with regard to registral direction and intervallic difference.

The principle of registral direction states that small intervals imply a following interval in the same registral direction (a small upward interval implies another upward interval and analogously for downward intervals), and large intervals imply a change in registral direction (a large upward interval implies a downward interval and analogously for downward intervals). Based on these two principles, melodic patterns or groups can be identified that either satisfy or violate the implication as predicted by the principles.

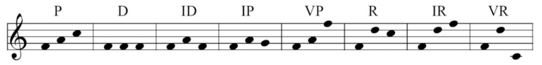


Fig. 3. Prototypical Narmour structures

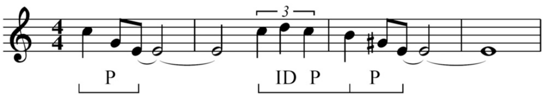


Fig. 4. Prototypical Narmour structures

Figure 3 shows prototypical Narmour structures. A note in a melody often belongs to more than one structure, i.e. a description of a melody as a sequence of Narmour structures consists of a list of overlapping structures. We parse each melody in the training data in order to automatically generate an implication/realization analysis. Figure 4 shows the analysis for a fragment of the Jazz standard *All of me*.

The resulting aligned sequences are cast into several inductive logic models (see II-C), one for note onset prediction, one for note duration prediction and one for note transition (legato) prediction. The models –capturing predictions for note onset and duration transformations– can be applied to a non-expressive input score, e.g. a MIDI or MusicXML description, to obtain an expressively enriched score that includes the corresponding note onset and duration transformations and expressive performance annotations used in synthesis. The legato predictor in particular is used to guide the sample selection process in the concatenative synthesizer, described in III-B.

C. Learning of expressive performance models

For building a computational model of a set of performances of a particular performer in a particular style, we use Tilde, a top-down decision tree induction algorithm [9]. Tilde can be considered as a first order logic extension of the C4.5 decision tree algorithm: instead of testing attribute values at the nodes of the tree, Tilde tests logical predicates. This provides the advantages of both propositional decision trees (i.e. efficiency and pruning techniques) and the use of first order logic (i.e. increased expressiveness). The increased expressiveness of first order logic not only provides a more elegant and efficient specification of the musical context of a note, but it provides a more accurate predictive model [10].

We apply the learning algorithm with target predicates `duration/3` and `energy/3`. (where `/n` at the end of the predicate name refers to the predicate arity, i.e. the number of arguments the predicate takes). Each target predicate corresponds to a particular type of transformation: `duration/3` refers to duration transformation and `energy/3` to energy transformation.

For each target predicate we use the complete training data specialized for the particular type of transformation as an example set, e.g. for `duration/3` we used the complete data set information on duration transformation (i.e. the performed duration transformation for each note in the data set). The arguments are the musical piece, the note in the piece and performed transformation.

We use (background) predicates to specify both note musical context and background information. The predicates we consider include `context/8`, `narmour/2`, `succ/2` and `member/3`. Predicate `context/8` specifies the local context of a note, i.e. its arguments are *(Note, Pitch, Dur, MetrStr, PrevPitch, PrevDur, NextPitch, NextDur)*: note identifier, note's nominal duration, duration of previous and following notes, extension of the pitch intervals between the note and the previous and following notes, and tempo at which the note is played. Predicate `narmour/2` specifies the Narmour groups to which the note belongs. Its arguments are the note identifier and a list of Narmour groups. Predicate `succ(X,Y)` means `Y` is the successor of `Y`, and Predicate `member(X,L)` means `X` is a member of list `L`. Note that `succ(X,Y)` also mean `X` is the predecessor of `Y`. The `succ(X,Y)` predicate allows the specification of arbitrary-size note-context by chaining a number of successive note (4).

$$succ(X_1, X_2), succ(X_2, X_3), \ldots, succ(X_{n-1}, X_n) \quad (4)$$

where X_i $(1 \leq i \leq n)$ is the note of interest.

III. Expressive performance synthesis

Synthesis of expressive performance in our case is the problem of rendering a score that has been previously enriched with expressivity annotations drawn from an inductive logic model by means of concatenative synthesis.

Speech synthesis research indicates, that one of the most important aspects to be handled by a convincing synthesizer are transitions between phones in addition to the phones themselves [11]. When applied to musical instrument synthesis this means that, depending on the different playing techniques supported by a particular instrument, inter-note transitions will play an important role in the final model. In our case of jazz saxophone, we concentrated on the most notable tongued vs. legato note transition differentiation: When a note is "tongued" on a reed instrument, the air flow between mouth cavity and mouthpiece is interrupted for a short period of time by the player putting his tongue slightly below the tip of the reed. When the interruption is released again and air continues to flow, the build-up in mouth-cavity pressure causes an increase in reed excitation of the air column in the instrument. The effect of tonguing can range from a very subtle and barely noticable alteration of the attack phase of a note to a very pronounced maximum in the energy envelope. Playing legato, on the other hand, means that the air pressure is held more or less constant over the course of a phrase and transitions in pitch to new notes are cause merely by opening or closing keys and thus shortening or lengthening the resonating air column.

As noted above, the approach taken here is to build separate models on the note level for predicting onset and energy transformations, as well as on the intra-note level, for predicting the type of transition to the next note.

A. Concatenative synthesis

The enriched, expressive score is used as input to a concatenative synthesizer [12], which, apart from depending on expressivity annotations, is completely independent from the rest of the system. The synthesizer reads an annotated input score in an extended WaveSurfer format [1].

The concatenation unit database is comprised of more than 3000 individual saxophone notes extracted from complete, expressively performed phrases. Notes are extracted and identified with their corresponding features as described in II-A.

B. Unit selection algorithm

The unit database is searched with a dynamic programming algorithm adapted from [13], [14] as described in [15] to find a sequence of database samples matching a given input score according to a cost function based on pitch, duration, timbre and musical context.

The algorithm starts by constructing a node-cost matrix with columns representing input score notes and rows denoting samples from the corpus, hereby associating a set of candidate samples $S\{t\}$ with each input score note at time t. In order to cut down the computational cost of the sample search per input score note, the sample database is divided into groups of similar samples by offline clustering, based on intra-note spectral features (currently spectral centroid, see II-A). From this clustering an additional model is created that can be applied to each input score note and annotates the note with a cluster number, which in turn determines the candidate list during sample search.

[1] Additional `ATTRIBUTE:VALUE` annotations are passed on the same line as arguments to a note, delimited by spaces

At each point in time t –i.e. for each input score note– the current path cost (5) is recursively calculated for each of the corresponding candidate samples $i \in S\{t\}$, based on the current node cost $\hat{C}_i(t)$ and the previous path cost $C_j(t-1)$. Finally, the optimal path is traced from the node corresponding to the last score note that has the minimal overall path cost to the beginning of the score.

$$C_i(t) = min_{j \in S\{t-1\}}[\hat{C}_i(t) + C_j(t-1)] \qquad (5)$$

The cost function (6) is a weighted sum of two separate cost functions: The transformation cost (7) is computed from feature distance costs $F_{Tj}(t,i)$ of the score note at time t and the sample candidate i, weighted by weights w_{tj}.

$$\hat{C}_i(t) = w_T \hat{C}_i^T(t) + w_C \hat{C}_i^C(t) \qquad (6)$$

$$\hat{C}_i^T(t) = \sqrt{\sum_j (w_{Tj} F_{Tj}(t,i))^2} \qquad (7)$$

The features costs used include pitch transformation cost, energy transformation cost, duration compression and expansion costs, and the transformation costs associated with the interval to the previous and the next note, respectively.

The concatenation cost (8) is composed of features based on the score note t, the sample database candidate note i and the currently accumulated best path p. The features currently used are path transition cost based on the legato descriptor, a duplicate feature penalizing reusing the same sample for different notes in the rendered phrase, a phrase membership cost rewarding the use of successive samples from the same database phrase, and a heuristic spectral continuity cost, rewarding a smooth spectral envelope for the resulting phrase based on the spectral centroid feature.

$$\hat{C}_i^C(t,p) = \sqrt{\sum_j (w_{Cj} F_{Cj}(t,i,p))^2} \qquad (8)$$

Once an optimal mapping from database notes to input score notes is found, the phrase segments corresponding to notes are transformed in time and pitch by means of spectral peak processing [16], [17] and concatenated in the frequency domain in order to yield a high fidelity rendering of the expressive input score. Special care is taken to preserve attack and release segments and in particular sample transitions by only time-stretching the sustain part of the amplitude envelope.

IV. Conclusions and future work

We have presented a concatenative synthesis system using dynamic programming for determining the best sequence of database samples based on a cost function derived from high-level musical features. An inductive logic model for predicting inter-note transitions in an input score was used to improve note transition quality during concatenative synthesis. Future work will concentrate on refining transition modeling, because informal listening evaluations suggest that this is the most critical area for improvements. Another important planned undertaking is the systematic construction of a comprehensive sample database dedicated to synthesis.

Acknowledgments

This work is supported by the Spanish PROSEMUS project (TIN2006-14932).

References

[1] G. Widmer and W. Goebl, "Computational models of expressive music performance: The state of the art," *Journal of New Music Research*, vol. 33, no. 3, pp. 203–216, 2004.

[2] J. Kaiser and R. Schafer, "On the use of the i0-sinh window for spectrum analysis," *Acoustics, Speech, and Signal Processing [see also IEEE Transactions on Signal Processing], IEEE Transactions on*, vol. 28, no. 1, pp. 105–107, Feb 1980.

[3] R. C. Maher and J. W. Beauchamp, "Fundamental frequency estimation of musical signals using a two-way mismatch procedure ," *Acoustical Society of America Journal*, vol. 95, pp. 2254–2263, Apr. 1994.

[4] R. Ramirez and A. Hazan, "Inducing a generative expressive performance model using a sequential-covering genetic algorithm," in *GECCO '07: Proceedings of the 9th annual conference on Genetic and evolutionary computation*. New York, NY, USA: ACM, 2007, pp. 2159–2166.

[5] A. Klapuri, "Sound onset detection by applying psychoacoustic knowledge," in *ICASSP '99: Proceedings of the Acoustics, Speech, and Signal Processing, 1999. on 1999 IEEE International Conference*. Washington, DC, USA: IEEE Computer Society, 1999, pp. 3089–3092.

[6] R. McNab, L. Smith, and I. Witten, "Signal processing for melody transcription," in *Proc. 1996 Australasian Computer Science Conference*, Melbourne, Australia, January 1996, pp. 301–307.

[7] E. Maestre and E. Gómez, "Automatic characterization of dynamics and articulation of monophonic expressive recordings," in *Proceedings of the AES 118th International Conference*, 2004.

[8] E. Narmour, *The Analysis and Cognition of Basic Melodic Structures*. Chicago and London: The University of Chicago Press, 1990.

[9] H. Blockeel, "Top-down induction of first order logical decision trees," Ph.D. dissertation, Department of Computer Science, Katholieke Universiteit Leuven, 1998.

[10] R. Ramirez, A. Hazan, E. Maestre, and X. Serra, *A Machine Learning Approach to Expressive Performance in Jazz Standards*. Springer, 2006.

[11] M. Beutnagel, A. Conkie, and A. Syrdal, "Diphone synthesis using unit selection," in *The 3rd ESCA/COCOSDA Workshop on Speech Synthesis*, Jenolan Caves, NSW, Australia, November 1998.

[12] D. Schwartz, "Data-driven concatenative sound synthesis," Ph.D. dissertation, University of Paris 6 – Pierre et Marie Curie, 2004.

[13] A. J. Viterbi, "Error bounds for convolutional codes and an asymptotically optimum decoding algorithm," in *IEEE Transactions on Information Theory*. IEEE, April 1967, vol. 13, no. 2, pp. 260–269.

[14] G. D. Forney, Jr., "The viterbi algorithm," in *Proceedings of the IEEE*, vol. 61, no. 3, March 1973, pp. 268–278.

[15] A. Hunt and A. Black, "Unit selection in a concatenative speech synthesis system using a large speech database," *Acoustics, Speech, and Signal Processing, 1996. ICASSP-96. Conference Proceedings., 1996 IEEE International Conference on*, vol. 1, pp. 373–376 vol. 1, May 1996.

[16] J. Laroche and M. Dolson, "Improved phase vocoder time-scale modification of audio," *IEEE Transactions on Speech and Audio Processing*, vol. 7, no. 3, pp. 323–332, May 1999.

[17] J. Bonada, "Automatic technique in frequency domain for near-lossless time-scale modification of audio," in *Proc. 2000 International Computer Music Conference*, 2000. [Online]. Available: citeseer.ist.psu.edu/529334.html

Designing and Synthesizing Delay-Based Digital Audio Effects using the CORDIS ANIMA Physical Modeling Formalism

Kontogeorgakopoulos Alexandros*, Cadoz Claude †

* ACROE-ICA laboratory INPG, Grenoble, France, Alexandros.Kontogeorgakopoulos@imag.fr
† ACROE-ICA laboratory INPG, Grenoble, France, Claude.Cadoz@imag.fr

Abstract — **Throughout this paper, several CORDIS-ANIMA physical models will be presented to offer an alternative synthesis of some classical delay-based digital audio effects: a delay model, two comb filter models, three flanger models and a sound spatialization model. Several of these realizations support a control scheme based on the "Physical Instrumental Interaction". Additionally they provide several sonic characteristics which do not appear in the original ones. Especially the flanger model, for certain parameter values may give a new digital audio effect between flanging and filtering**.

I. Introduction

A variety of digital audio effects (DAFx) make use of time delays. The echo, the comb filter, the flanger, the chorus and the reverb for example use as building block the time delay [1][2]. An evident digital synthesis-realization of the time delay is the digital delay line. Generally, the digital delay line and the unit delay is one of the few basic building blocks used in almost every audio digital signal processing algorithm.

However, when designing musical sound modification and sound synthesis units, a very important criterion, apart from the algorithm itself, is their control. In our digital audio effect designs, the " Physical Instrumental Interaction " - the type of physical interaction which a musician establishes with a real musical instrument- is fundamental [3][4][5]. Hence the synthesis of sound processing algorithms using structures that offer this type of interaction is the base of this research.

In the present article, after taking a brief look in the CA physical modelling and simulation system and its electrical analog, a number of digital audio effects models will be presented. A schematic block diagram will follow each model. These diagrams are combinations of CA networks with classical digital signal processing block diagrams; they describe in detail the sound modification algorithms.

II. Cordis-Anima System

A. CA network

In this essay, the proposed audio effects are actually computer models of physical objects. Moreover, several gesture models complete the "alphabet" from which the modification algorithms are designed. Generally, our toolbox contains a set of elementary virtual mechanical components –the CA modules (figure 1) - and exceptionally simple digital signal processing building blocks such as adders and multipliers.

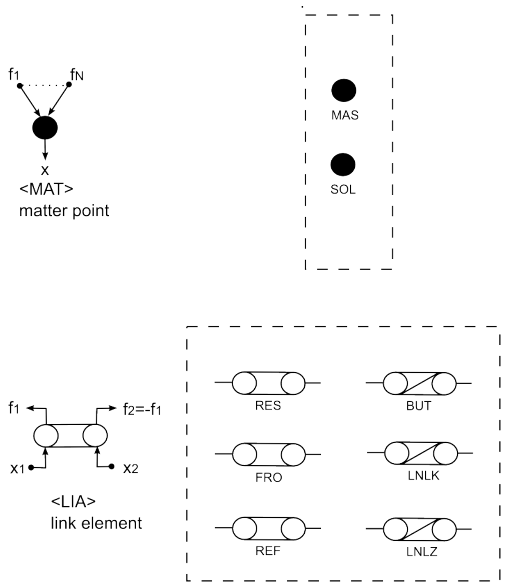


Fig. 1. CA modules

Each DAFx model is represented by a plane topological network whose nodes are the punctual matter elements <MAT> and links are the physical interaction elements <LIA> according to the CA formalism [6]. The simulation space used is limited to one dimension. Forces and displacements are projected on a single axis, perpendicular to the network plane. In tables I and II we depict the algorithms of CA modules. Table I provides the linear CA modules and table I the nonlinear CA modules.

These algorithms can easily take the form of ordinary signal processing block diagrams. However we prefer the CA network representation since its use makes the physical constitution of models easily perceptible and detectible.

In some models it was inevitable to enrich the CA networks with simple digital signal processing modules as adders, multipliers, switchers and filters. However the use of "externals" has been carried out with care in order to preserve the energetic passivity on certain parts of the models that was necessary.

TABLE I.
CA LINEAR ALGORITHMS

● MAS	$x(n) = 2x(n-1) - x(n-2) + \frac{1}{M}\sum_{i}^{N} f^{i}(n-1)$
● SOL	$x(n) = c$
RES	$f^{ij}_{RES}(n) = K_{ij}\Delta x$ $\Delta x = [x_i(n) - x_j(n)]$
FRO	$f^{ij}_{FRO}(n) = Z_{ij}\Delta v$ $\Delta v = [x_i(n) - x_i(n-1) - x_j(n) + x_j(n-1)]$
REF	$f^{ij}_{REF}(n) = f^{ij}_{RES}(n) + f^{ij}_{FRO}(n)$

TABLE II.
CA NONLINEAR ALGORITHMS

BUT	$f^{ij}_{BUT}(n) = \begin{cases} f^{ij}_{REF}(n) & x_i(n) - x_j(n) \leq S \\ 0 & x_i(n) - x_j(n) > S \end{cases}$
LNLK	$f^{ij}_{LNLK}(n) = \text{interpolate}(\Delta x, \mathrm{T})$ lookup table $\mathrm{T} : (f_r, \Delta x_r) \quad 2 \leq r \leq 20$ interpolation types : linear, cubic, splines
LNLZ	$f^{ij}_{LNLZ}(n) = \text{interpolate}(\Delta v, \mathrm{T})$ lookup table $\mathrm{T} : (f_r, \Delta v_r) \quad 2 \leq r \leq 20$ interpolation types : linear, cubic, splines

B. CA Electrical Analogs (Krirchhoff) : one-ports and 2-ports

An electrical analogous circuit of a mechanical system is an electrical circuit in which currents/voltages are analogous to velocities/forces in the mechanical system. If voltage is the analog of force and current is analog of velocity the circuit is called impedance analogous. In a similar way if voltage is the analog of velocity and current is the analog of force the circuit is called mobility analogous.

CA networks can be seen as a discrete time approximation of a subclass of mass-spring systems. Therefore electrical circuits may represent them easily. In practice to pass form a CA network to Kirchhoff network we transform in a first step the CA network into a mass-spring network. Then using the classical electro-mechanic analogies we obtain the Kirchhoff circuit [7].

The arrows represent the transition from one system formalism to another. We use a simple arrow symbol for the certain transition, a dotted arrow for the unsure transition and a marked arrow for the approximate transition. Table III gives the analogies between CA and electrical circuits variables and parameters.

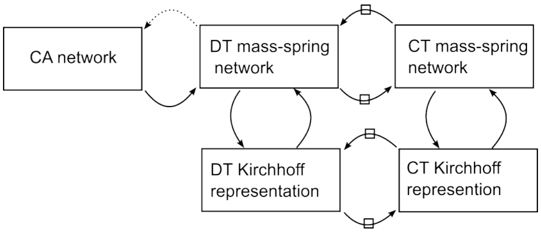


Fig. 2. From CA networks to discrete-time (DT) and to continuous-time (CT) Kirchhoff representation

TABLE III.
CA ELECTRICAL ANALOGS

Impedance Analog	Mobility Analog
$f \rightarrow \upsilon$	$f \rightarrow i$
$x \rightarrow q$	$x \rightarrow \varphi$
$M \rightarrow L$	$M \rightarrow C$
$K \rightarrow 1/C$	$K \rightarrow 1/L$
$Z \rightarrow R$	$Z \rightarrow 1/R$

III. DIGITAL AUDIO EFFECTS MODELS

A. CA Delay Model

A delay simply delays the input audio signal by an amount of time. For fractional sample delay lengths, interpolation techniques are used such as linear and allpass interpolation algorithms [1][2].

Synthesizing a delay line with the mass-interaction physical modeling scheme of CA is neither straightforward nor computationally effective. On the other hand, its algorithmic structure can be interesting as it offers a mentally strong physical metaphor and permits directly a control based on the ”Physical Instrumental Interaction”.

From figure 2 we observe that often we can pass from a continuous-time Kirchhoff network to a CA one. Luckily an electrical transmission line, which may be considered as an analog delay line, can be transformed to CA structure. Hence, in CA system, a digital delay line takes the form of virtual string terminated by its characteristic impedance.

In figure 3 we depict the network of an electrical delay line and its CA realization. The impedance analog network has been used. The stiffness parameter K of the model controls the time delay. Analytic expressions of the time delay as a function of the model will be presented later on.

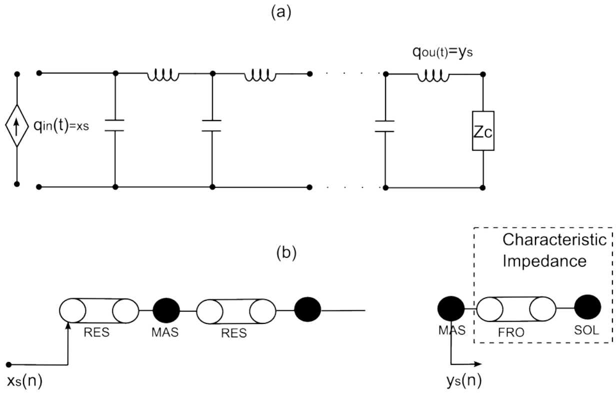


Fig. 3. (a) electric transmission line (b) CA delay model (xs: input sound, ys: output sound)

We can compute the characteristic impedance by expressing and decomposing the CA model into two-ports. An approximate analytic expression of the time delay as a function of the model parameters can be computed as well by the same decomposition. Figure 4 illustrates an elementary CA two-port used. Equation (1) expresses mathematically its input/output terminal relations.

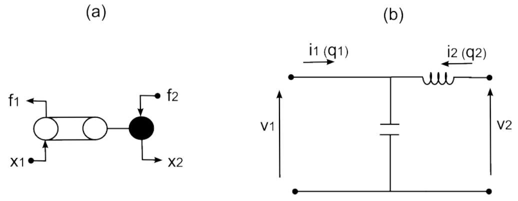


Fig. 4. From CA networks to discrete-time (DT) and to continuous-time (CT) Kirchhoff representation

$$\begin{pmatrix} f_1 \\ f_2 \end{pmatrix} = \begin{pmatrix} K & -K \\ K & -K - M\dfrac{1-2z^{-1}+z^{-2}}{z^{-1}} \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} \qquad (1)$$

The concept of applying the of two-ports representation to the digital signal processing domain is not new as it has been introduced earlier in a similar way by Mitra [8]. However in the CA formalism it has never been used before.

The terms of the matrix that appears in the equation (1) are called impedance parameters zij [9]. The characteristic impedance Zc and the delay time D are functions of those parameters [10]. Equations (2) and (3) give their analytic mathematical expressions (we express the impedance parameters by their Fourier-transform). The total time delay for a N-MAS CA delay line is given by the equation (4).

$$Zc = K - K\frac{2Ke^{-j\omega} + M(1-2e^{-j\omega}+e^{-2j\omega})}{2Ke^{-j\omega}} \pm K\sqrt{\left(\frac{2Ke^{-j\omega} + M(1-2e^{-j\omega}+e^{-2j\omega})}{2Ke^{-j\omega}}\right)^2 - 1} \qquad (2)$$

$$D = \frac{d\cos^{-1}(\lambda)}{d\omega} \text{ with } \lambda = \frac{z_{22}-z_{11}}{2z_{12}} \overset{(1)}{\Rightarrow}$$

$$D = \frac{d\cos^{-1}\left(\dfrac{2Ke^{-j\omega} + M(1-2e^{-j\omega}+e^{-2j\omega})}{2Ke^{-j\omega}}\right)}{d\omega} \qquad (3)$$

$$D_{total} = ND \overset{(3)}{\Rightarrow}$$

$$D_{total} = N\frac{d\cos^{-1}\left(\dfrac{2Ke^{-j\omega} + M(1-2e^{-j\omega}+e^{-2j\omega})}{2Ke^{-j\omega}}\right)}{d\omega} \qquad (4)$$

We observe from the last expression (equation (4)) and we also perceive from the audio outputs that the model suffers from dispersion. It is remarkable that for M=K=1 certain values the CA delay line synthesizes precisely the time delay without phase distortion or undesired filtering. The Zc and D in this case are:

-

$$Z_c = 1 - z^{-1} \qquad (5)$$

$$D = 1 \text{ and } D_{total} = N \qquad (6)$$

The characteristic impedance expressed by the equation (5) can be synthesized in CA by a FRO module with Z=1 attached to a SOL module.

Instead of the precise derived equations (2)-(4) we may use their approximation in the continuous-time domain as given by the continuous-time electrical network. Using the results from the electrical transmission lines [10] and the CA-electrical analogs we get the following helpful approximations:

$$Z_c = \sqrt{MK} \text{ for } \omega^2 \frac{M}{K} << 4 \qquad (7)$$

$$D = \sqrt{\frac{M}{K}} \text{ and } D_{total} = N\sqrt{\frac{M}{K}} \qquad (8)$$

B. CA Comb Filter Models

A comb filter is produced when a slightly delayed audio signal is mixed with the original one [1][2][11]. When the delayed version is fed back to the delay line input we have an IIR comb filter. Otherwise we get an FIR comb filter. Both topologies give a large number of notches in the spectrum of the input signal.

Two models that synthesize this classical digital audio effect are depicted in figure 2. The first one uses the = delay model. A similar effect to the IIR one is experience naturally inside an acoustical cylinder when a sound circulates inside it: the successive reflections at both ends of a cylinder modify the signal approximately as a IIR comb filter. It is not difficult to simulate this phenomenon with a CA string model (second model of figure 2). The resulted effect is perceived as a natural resonator. Two important differences from the signal processing model are 1. the notches do not cover the whole spectrum (the number of notches are defined by the number of masses) and 2. the dispersion is inevitable in contrast to the first one where it is inexistent.

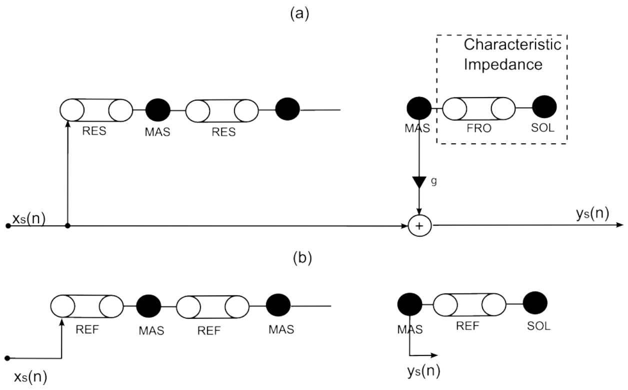


Fig. 5 (a) CA comb filter model using the delay model (b) CA comb filter model using a CA string (xs: input sound, ys: output sound)

C. CA Flanger Models

A flanger can be easily implemented using variable length delay lines [1][2][12][13][14][15]. Basically it is a comb filter where the delay line length is slightly changed. Hence we may use the previous CA comb filter models where the parameter K that defines the time delay approximately according to equation (8) is altered periodically (figure 6a). For wide amplitude modulation values of the K parameter, a new effect is obtained between dynamic filtering and flanging.

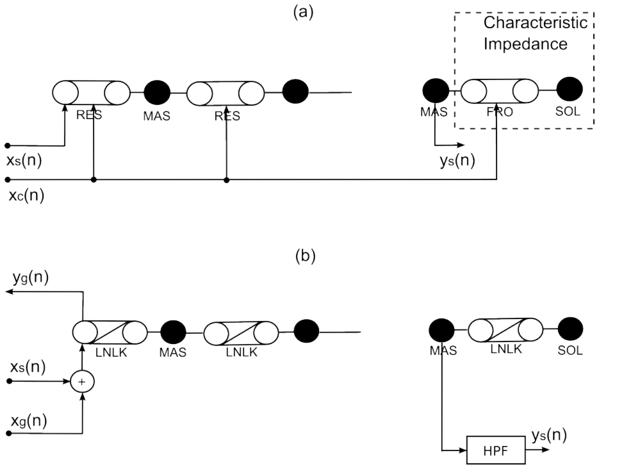


Fig. 6 (a) CA flanger model using a comb filter model (b) CA flanger model using a nonlinear CA string (xs: input sound, ys: output sound, xc: control input, xg: gesture input, yg: gesture output)

A more physical approach in the context of "Physical Instrumental Interaction" is to use a variation of the second CA comb filter model where the linear <RES> modules are exchanged by non-linear <LNLK> modules. The designed non-linearity characterizes the time delay. A gesture stressing the physical model biases it and consequently determines the linear regions of the system. Therefore this gesture affects the time delay of the comb filter structure.

Another flanger model will be presented in the next section. It has been preferred to be described in that part of the paper because of the adopted design approach.

D. CA Spatialization Model + Flanger Model

The CA networks have an inherent spatiality due to their topology. The sound can be picked-up from every elementary CA basic module output. Figure 7 represents a simple CA flanger model with two outputs. Due to the relative time delay between the two output nodes in the network, we obtain a spatial image of the sound source (the interaural time differences (ITD) are a strong cue that the human auditory system uses to estimate the apparent direction of a sound source [1]). Hence the geometrical spatial characteristics (a more accurate term would be topological) of CA models are quite related to the spatial sound characteristics of the outputs.

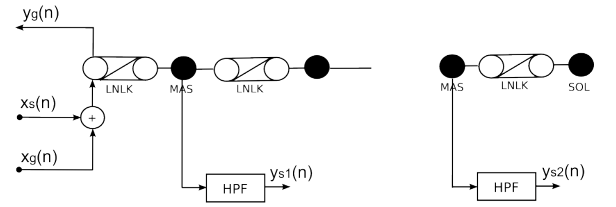


Fig. 7 (a) CA spatilization model (xs: input sound, ys: output sound, xg: gesture input, yg: gesture output)

The analysis of the chapter III.A can be applied in order to choose the proper CA nodes to obtain a desired spatial image of the sound source. It is clear that the spatial discretization quantizes the spatial trajectories. The interaural intensity differences (IID) can be used as well to improve the spatial image.

The use of movable pick-up points gives the opportunity to attain dynamical effects. A similar idea has been applied earlier to digital waveguides [16]. The CA model of figure 8 is a type of flanger. Each pick up point determines the partials reinforced by the string topology. If we place the pick-up point at a position 1/m across the sting length, the partials whose number is m will be canceled.

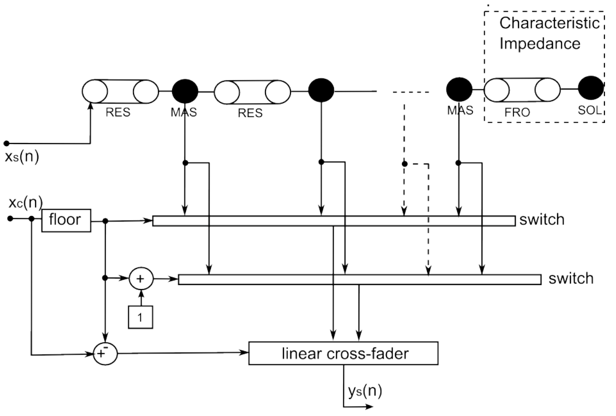


Fig. 8 CA flanger model using a pick-up point modulation (xs: input sound, ys: output sound, xc: control input)

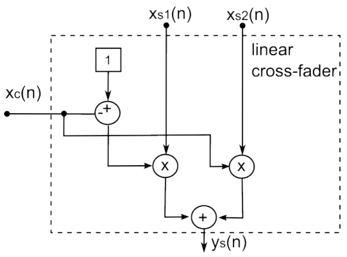


Fig. 9 Linear cross-fader block diagram

IV. FURURE WORKS AND CONCLUSIONS

All of the above models were simulated and run in differ time. Several other simple models were designed to simulate the physical mechanical gestures used to control and interact with them. In the near future all of them and several others used as well for sound processing will be implemented in real time and controlled by force feedback haptic interfaces as the TGR developed in ACROE laboratory [17].

The aim of this general research is to develop new tools for sound transformation mainly used for musical purposes that preserve the important natural instrumental relation found in acoustical musical instruments. We believe that through this physical dynamic control of the audio effect process, a virtuosity will emerge that will contribute to the quality and the finesse of sound transformation.

REFERENCES

[1] U. Zoelzer (Edt), *Digital Audio Effects*, John Wiley & Sons Ltd, 2002

[2] J. O. Smith, *Physical Audio Signal Processing For Virtual Music Instruments and Digital Audio Effects, December 2005 DRAFT*

[3] C. Cadoz, M. M. Wanderley, "Gesture-Music", in *Trends in Gestural Control of Music*, M. M. Wanderley and M. Battier, eds, ©2000, Ircam – Centre Pompidou, pp. 71-94 , 2000

[4] A. Kontogeorgakopoulos, C. Cadoz, "Filtering Within the Framework of the Mass-Interaction Physical Modelling and of Haptic Gestural Interaction", *Proceedings of Digital Audio Effects DAFX07*, Bordeaux, France, 2007

[5] A. Kontogeorgakopoulos and C. Cadoz, "Amplitude Modification Algorithms in the Framework of Physical Modeling and of Haptic Gestural Interaction", *in Proc. Audio Engineering Society Convention AES124*, Amsterdam, Netherlands, 2008

[6] C. Cadoz, A. Lucian and J.L Florens, "CORDIS-ANIMA: A modelling and Simulation System for Sound and Image Synthesis – The General Formalism", Computer Music Journal, 17(4*), 1993*

[7] A. Kontogeorgakopoulos, C. Cadoz, "CORDIS-ANIMA Physical Modeling and Simulation System Analysis", *Proceedings of SMC07*, Lefkada 2007

[8] S. K. Mitra, *Digital Signal Processing: A Computer-Based Approach, McGraw-Hill, second edition,2001*

[9] J. W. Nilsson, S. A. Riedel, *Electric Circuits*, Pearson Education, 1996

[10] Y. Rocard, *Dynamique générale des vibrations*, Masson et Cie, 1971

[11] P. Dutilleux, "Filters, delays, modulations and demodulations: A tutorial" *Proceedings of Digital Audio Effects DAFX98, Barcelona, Spain, 1998*

[12] J. O. Smith, "An all-pass approach to digital phasing and flanging", *Proceedings of the 1984 International Computer Music Conference*, Paris, France, 1984

[13] J. Dattorro, "Effect design part 2: delay line modulation and chorus", *JAES* 45(10), 1997

[14] S. Disch, U. Zolzer, "Modulation and delay line based digital audio effects", *Proceedings of the 200 Digital Audio Effects Conference*, Trondheim, Norway, 1999

[15] A. Huovilainen, "Enhanced digital models for digital modulation effects", *Proceedings of the 2005 Digital Audio Effects Conference*, Madrid, Spain, 2005

[16] S. A. Van Duyne, and J. O. Smith, "Implementation of a variable pick-up point on a waveguide string model with FM/AM applications", *In Proc. Intl. Computer Music Conf*, San Jose, pp. 154–157, 1992

[17] C. Cadoz, A. Luciani and J.-L. Florens, "Responsive Input Devices and Sound Synthesis by Simulation of Instrumental Mechanisms : The CORDIS System", *Computer Music Journal* 8(3), 1984 – reprint in C. Roads (Edt), *The Music Machine*, The MIT Press, 1988

Extending voice-driven synthesis to audio mosaicing

Jordi Janer, Maarten de Boer
Music Technology Group, Universitat Pompeu Fabra, Barcelona

***Abstract*—This paper presents a system for controlling audio mosaicing with a voice signal, which can be interpreted as a further step in voice-driven sound synthesis. Compared to voice-driven instrumental synthesis, it increases the variety in the synthesized timbre. Also, it provides a more direct interface for audio mosaicing applications, where the performer voice controls rhythmic, tonal and timbre properties of the output sound. In a first step, voice signal is segmented into syllables, extracting a set of acoustic features for each segment. In the concatenative synthesis process, the voice acoustic features (target) are used to retrieve the most similar segment from the corpus of audio sources. We implemented a system working in pseudo-realtime, which analyzes voice input and sends control messages to the concatenative synthesis module. Additionally, this work raises questions to be further explored about mapping the input voice timbre space onto the audio sources timbre space.**

I. Introduction

State of the art synthesizers are able to generate realistic sounds, using physical models or advanced sample-based techniques. In addition, feature-driven synthesis of audio material is a recently emerging field. A particularly recognizable instance of this field is Audio Mosaicing, the practice of automatically assembling micro-segments of songs, or other audio, to match a pre-determined source. At the same time, major challenges in current digital musical instruments are on the control side. Since the advent of MIDI, a wide variety of musical interfaces (musical controllers) have been proposed to control synthesizers. In this context, the use of the voice to control sound synthesis represents an interesting path for improving music interaction. In this paper, we aim to extend voice-driven synthesis from the control of instrumental sound synthesis to the control of audio mosaicing. Voice input controls the rhythmic, tonal and timbre properties of the output sound, which combined with a loop based mechanism becomes an appropriate system for live performing.

A. Related work

Regarding the use of voice in music interaction, audio-driven synthesis [1] uses features from an input audio signal, usually from another instrument, to control a synthesis process. When using the voice as input signal, the front-end has been also referred to as singing-driven interface [2]. For the latter case, previous research addressed the characteristics of the singing voice in instrument imitation, highlighting the role of phonetics in terms of timbre and musical articulation. The present work seeks to exploit the timbre possibilities of the voice input to find similar audio micro-segments from other audio sources. This differs from other voice-related approaches in the area of Music Information Retrieval such as query-by-humming (QbH) that retrieves a song from voice melody [3], or query-by-beatboxing that retrieves a drum loop from voice timbre sequence [4]. Compared to the first, QbH systems apply a melody transcription of the voice input and the searches for is done in the symbolic domain (usually MIDI) without taking into account timbre information. Compared to the second, which searches for existing drum loops based on timbre and rhythmic similarity, our search unit is not a loop but a micro-segment and the generated sound is not limited to any pre-recorded sound loop. Furthermore, our system uses tonal information of the voice input when retrieving non-percussive sounds, e.g instrumental chords or single notes. Yet another approach [11] uses voice input timbre to train a 3-class classifier during the preparation phase and, in the performance phase, the output of the classifier is used to trigger three different percussive sounds.

Regarding audio mosaicing, it is a concatenative sound synthesis technique that has gained interest in the recent years [5], [6]. Audio or musical mosaicing aims at automatically, with or without user intervention to generate sequences of sound samples by specifying high-level properties of the sequence to generate. These properties are typically translated into constraints holding on acoustic features of the samples. The field also expands to include live, performance-oriented systems. One similar approach is found in *sCrAmBlEd?HaCkZ!* [7], a system that uses a speech input to generate a sequence of similar sounds. The principal differences compared to our system are that in our case the voice input drives also tonal properties (”chord”) of the synthesized output segment. We propose to work on a loop-based synthesis, where tempo and the loop length (number of segments) are modifiable. Also, we can build the vocal target loop by layering several voice input sequentially. e.g beatboxing a drum-line, adding later a bass-line, and adding other sounds on top. Summarizing, the objective of this work is to provide a live input control to an existing audio mosaicing system [10], and in particular addressing the necessary components when using the singing voice as input signal.

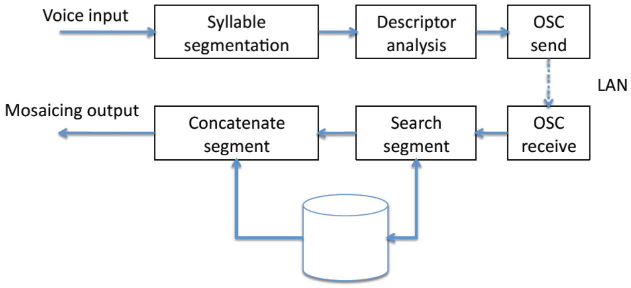


Fig. 1. Block diagram of the proposed system.

Feature	*Characteristic*
audioCentroid	*impulsiveness*
energy	loudness
flatness	noisiness/harmonicity
hpcp	tonal description
mfcc	timbre
spectralCentroid	brightness

TABLE I
LIST OF THE ACOUSTIC FEATURES EXTRACTED.

II. System description

A. Overview

The proposed system is composed of two modules: the interface and the feature-based synthesis engine. Regarding the musical output, the synthesized sound is loop-based, where the audio content resembles the characteristics of the vocal input loop(target), which can be altered on the fly by the user with additional control parameters. As shown in figure 1, the vocal input is segmented into syllables. For each voice segment, a vector of target features is generated, which is used to find the most similar audio segment in the selected audio sources. The corpus of audio sources are song excerpts with a duration of several seconds and which have been segmented using a general onset detector.

B. Voice description

Voice input has a known tempo and the number of segments that conform the loop is also known. In a first step, the voice signal is segmented using an algorithm specifically designed for instrument imitation signals [2], which performs better than general purpose onset detection algorithms. The latter gives a lot of false positives since it is not adapted to voice signals. Our segmentation algorithm relies on heuristic rules that primarily looks at phonetic variations based on musical functions (attack, sustain, release, etc.).

The second step is to extract acoustic features. In voice-driven instrumental sound synthesis, instantaneous voice features were extracted in short-time frames, capturing the time evolution of pitch, energy, formants and degree of phonation (breathiness). In contrast, in the present approach, we work at a segment level, computing one vector of features per segment of a duration of one beat. Table I lists the acoustic features, including tonal descriptors (HPCP) [8], rhythm (audio centroid) and timbre (MFCC). Actually, the vector of acoustic features for one segment is the mean value of the instantaneous frame values computed using a hop-size of 512 samples and a window size of 2048 samples at a sampling rate of 44100 Hz.

C. Feature-driven synthesis

If we go back to the general description of audio mosaicing, we can look at the individual steps, and identify four processes: target selection, source selection, unit selection and concatenation. Essentially, the system operates by looping the selected target and concatenating segments from the selected audio sources that best match the target. In our case, the selected target is a vocal signal segmented into syllables. The audio sources consist of song excerpts, which are previously analyzed and segmented using a generic onset detector. Each audio source i consists of a sequence of audio segments (with subindex j) from which a vector with the same set of acoustic features y_i is extracted (see table I).

The user selects a reduced number of audio sources (usually a dozen of song excerpts or *loops*) that will constitute the corpus. The user is then able to interactively change the selected source material, as well as to interactively augment or diminish the presence of a particular source on the fly. This concept is further described in [10]. The unit selection process retrieves for each target segment a list of similar units in the audio sources using a distance measure. The distance measure compares two vectors (a voice target x and an audio source y_{ij}) of acoustic features, where the presence of a given audio source is controlled by applying weights in the distance measure. In the equation 1, d_{ij} is the euclidean distance between target x and source y_{ij}, where w_i is the weight of the ith audio source, j is the segment index in an audio source i, and M is the number of acoustic features.

$$d_{ij} = w_i \sqrt{\sum_{m=1}^{M} (x_m - y_{ij,m})^2} \qquad (1)$$

Finally, the concatenation process includes layering and randomization of source segments simultaneously in order to produce a richer sound.

III. Prototype implementation

The implemented voice-driven interface analyzes the voice input signal and sends the target features x to the concatenative synthesis engine. In a typical work-flow, the user sets the tempo and the number of steps in the loop buffer. Then, he records the vocal input loop, which is stored in an internal buffer. The user can layer several vocal takes (e.g. imitating drums, bass line), thus creating a richer target sound.

Next, the system analyzes the internal buffer, extracting the acoustic features x. Finally, it sends for each target segment its acoustic features in a synchronous way to the synthesis engine. The latter is in charge of retrieving the

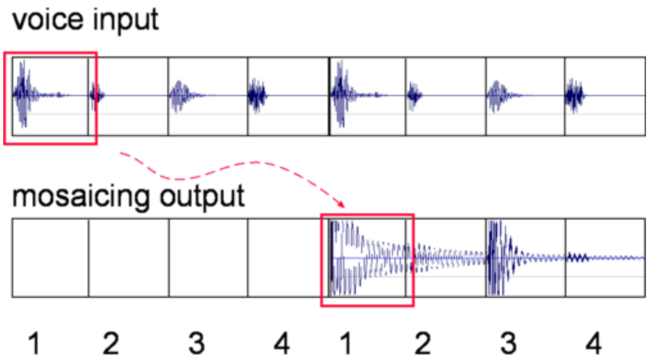
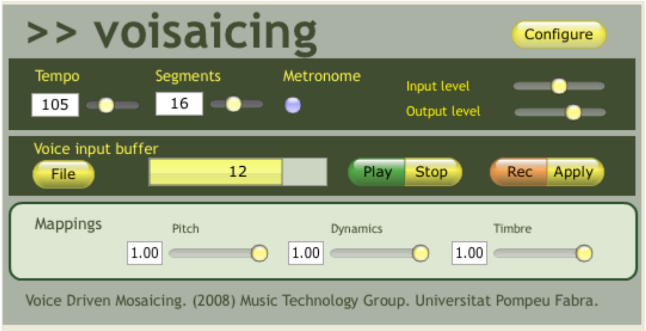


Fig. 2. Timeline of the pseudo-realtime work-flow with a loop length of 4 segments. The user sings a complete voice input loop before the target features are used in the mosaicing output.



Tempo 105
Segments 16
Metronome
Input level
Output level
Voice input buffer 12
Mappings
Pitch 1.00
Dynamics 1.00
Timbre 1.00

Voice Driven Mosaicing. (2008) Music Technology Group. Universitat Pompeu Fabra.

Fig. 3. Screen-shot of the interface module, implemented as a VST plugin.

nearest segment from the audio sources and concatenate the output stream.

Due to implementation considerations, the voice-driven interface and the concatenative synthesis engine are two separate processes that communicate each other through OpenSoundControl[1]. The interface module is a VST plug-in [2] working in pseudo-realtime, i.e. the voice input affects the output with a delay of one loop duration. Figure 2 shows the time evolution of the process, in this case with a loop length of 4 segments.

IV. Discussion on timbre mapping

First tests with the implemented system shows that the user is able to control the timbre of the synthesis output. However, in order to improve the sense of control, we suggest to study the use of mapping functions from voice features to audio source features for the retrieval.

Intuitively, the sonic spaces of the audio sources and the vocal imitation are different, so that a mapping function might be needed. This mapping function should allow to retrieve any sound in the corpus with a vocal input, thus mapping voice timbre space onto a larger sonic space. In figure 4, we represent the sonic space of vocal sounds (left) and the sonic space of the audio sources corpus (right), which is larger. We propose to learn the mapping functions by imitating a few examples and using statistical methods to derive the mapping functions.

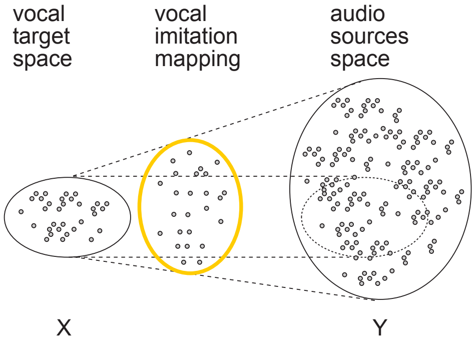


Fig. 4. Sonic spaces. Input voice target space(X), imitation mapping and output sonic space of the audio sources corpus (Y).

A. *Comparing timbre spaces*

In a preliminary experiment, we collected a corpus of 10 loops (from different musical genres) and the corresponding vocal imitations. The original audio loop and its imitation were segmented and aligned. Each subset consists of 144 short segments. Our goal is to study the timbre space of the voice input compared to the space of audio sources. A priori, the variance in the voice imitation features subset is likely to be lower than in the imitated audio sources subset, since audio source loops will contain any musical sound (including voice), and not inversely. As an initial test, we compute the Principal Component Analysis of both subsets separately using 13 Mel-Frequency Cepstrum Coefficients (MFCC) as data, normalized in a range $[0..1]$ over the complete set (voice imitation and audio sources). MFCC vector data is the mean value of the instantaneous MFCC values within a segment. Then, we project both subsets on the corresponding first two principal components. The variance explained by the first two PCA component are of 63.11% for the voice set, and 66.25% for the audio set. In figure 5, one can observe the projection of the two subsets, audio sources and voice imitation, where each segment corresponds to a *diamond* in the plot. Data is generated using the built-in Matlab function in *princomp*.

One can observe from the plots that the projection of the voice subset is more localized than the audio sources subset. It might indicate that the timbre variance of the voice segments is lower than the variance of the audio sources' timbre. However, we have to stress that this experiment uses a small amount of data, where 10 audio loops were imitated by a single subject.

B. *Mapping strategies*

In order to derive valid mapping functions, we should collect enough examples of audio loops and vocal imitation by several users. Then, to learn the mapping functions, we can used supervised training methods, where each vocal imitation segment is aligned with its corresponding imitated audio segment. Ideally, by building a sufficient large corpus of imitations for training, one can use statistical methods (e.g. gaussian mixtures) to model both source and target corpus and then find

[1] http://www.opensoundcontrol.org

[2] VST is trademark of Steinberg GmbH.

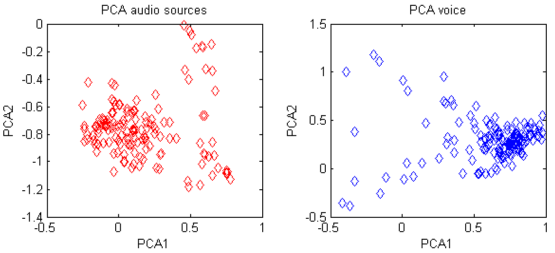


Fig. 5. Principal component analysis (PCA) of audio source (*left*) and voice imitation (*right*). Input data are vectors with 13 MFCC, which are previously normalized and projected on the corresponding first two PCA components.

the corresponding mapping functions using the models. This is similar to approaches found in voice conversion applications [12], where the timbre of the voice source has to be transformed to resemble the timbre of a target voice. Another strategy is to model non-linear mapping functions with neural networks. Alternatively, in cases where the training corpus is small, one possibility is to approximate the mapping functions with a linear regression of a reduced set of examples.

C. *User-specific mappings*

An additional issue concerning the mapping function is to implement a user-adapted system. We have two options: either to build user-dependent mapping functions, or build a general user-independent functions. In practice, we cannot assume the every individual imitate a sound in the same manner with his voice. Therefore, it seems more convenient to allow the system to learn the mapping functions for every user.

V. Conclusions

With the proposed system, we provide vocal control capabilities to one the synthesis techniques that has gained more interest in the recent years, audio mosaicing. Compared to voice-driven instrumental sound synthesis, this approach exploits in a higher degree the timbre possibilities of the human voice. At the same time, it offers a more direct way to interact with audio mosaicing, which is usually driven by graphical interfaces. Finally, this research has arisen questions about the mapping strategies between two different sonic spaces: input voice space and output audio source space. Audio examples demonstrating the achieved results can be found online[3].

Acknowledgment

This research has been partially funded by the IST project SALERO, FP6- 027122. Authors would like to thank to Graham Coleman, Perfecto Herrera and Martin Haro for the discussions.

References

[1] Jehan, T. and Schoner, B. (2001). 'An audio-driven, spectral analysis-based, perceptually meaningful timbre synthesizer'. In 110th Conv. Audio Eng. Soc., Amsterdam, Netherland.

[2] Janer, J. (2008), 'Singing-driven Interfaces for Sound Synthesziers', PhD Thesis Universitat Pompeu Fabra.

[3] Lesaffre, M., et al.(2003). 'The MAMI query-by-voice experiment: Collecting and annotating vocal queries for music information retrieval'. In Proceedings of the ISMIR 2003, 4th Int. Conf. on Music Information Retrieval, Baltimore.

[4] Kapur, A., Benning, M., and Tzanetakis, G. (2004). 'Query-by-beat-boxing: Music retrieval for the dj'. In ISMIR-2004.

[5] Zils, A. and Pachet, F. (2001). 'Musical Mosaicing'. In Proc. of the COST-G6 Workshop on Digital Audio Effects (DAFx-01), Limerick.

[6] Schwarz, D. (2005). 'Current Research in Concatenative Sound Synthesis'. Proceedings of the International Computer Music Conference (ICMC).

[7] http://www.popmodernism.org/scrambledhackz/

[8] Gómez, E. (2006). 'Tonal Description of Music Audio Signals'. PhD thesis, Universitat Pompeu Fabra.

[9] Peeters, G. (2003). 'A large set of audio features for sound description (similarity and classification) in the cuidado project'. IRCAM.

[10] Fujishima, T. et al. (2008). 'Music-piece processing apparatus and method', United States Patent 20080115658, Yamaha Corporation.

[11] Hazan, A. (2005). 'Billaboop real-time voice-driven drum generator'.Proceedings of the Digital Audio Effects Conference DAFX'05, Madrid.

[12] Dutoit, T. et al. (2007). 'Towards a voice conversion system based on frame selection', IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP).

[3] http://www.mtg.upf.edu/~jjaner/presentations/smc08

Organized by

DEGEM

ISBN 978-3-7983-2094-9

Vertrieb/
Publisher

Universitätsverlag der TU Berlin
Universitätsbibliothek
Fasanenstr. 88 (im VOLKSWAGEN-Haus)
10623 Berlin
GERMANY

Tel. (030)314-76131; Fax. (030)314-76133
E-Mail: publikationen@ub.tu-berlin.de
http://www.ub.tu-berlin.de/